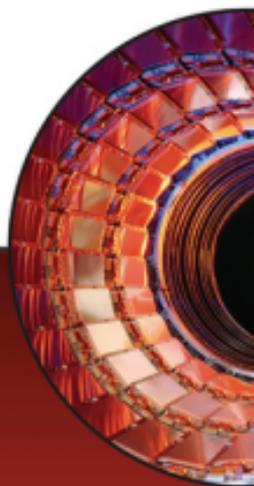


OXFORD MASTER SERIES IN PARTICLE PHYSICS,
ASTROPHYSICS, AND COSMOLOGY

Particle Physics in the LHC Era

G. Barr
R. Devenish
R. Walczak
T. Weidberg



OXFORD

series in condensed physics
in materials

PARTICLE PHYSICS IN THE LHC ERA

OXFORD MASTER SERIES IN PHYSICS

The Oxford Master Series is designed for final year undergraduate and beginning graduate students in physics and related disciplines. It has been driven by a perceived gap in the literature today. While basic undergraduate physics texts often show little or no connection with the huge explosion of research over the last two decades, more advanced and specialized texts tend to be rather daunting for students. In this series, all topics and their consequences are treated at a simple level, while pointers to recent developments are provided at various stages. The emphasis is on clear physical principles like symmetry, quantum mechanics, and electromagnetism which underlie the whole of physics. At the same time, the subjects are related to real measurements and to the experimental techniques and devices currently used by physicists in academe and industry. Books in this series are written as course books, and include ample tutorial material, examples, illustrations, revision points, and problem sets. They can likewise be used as preparation for students starting a doctorate in physics and related fields, or for recent graduates starting research in one of these fields in industry.

CONDENSED MATTER PHYSICS

1. M.T. Dove: *Structure and dynamics: an atomic view of materials*
2. J. Singleton: *Band theory and electronic properties of solids*
3. A.M. Fox: *Optical properties of solids, second edition*
4. S.J. Blundell: *Magnetism in condensed matter*
5. J.F. Annett: *Superconductivity, superfluids, and condensates*
6. R.A.L. Jones: *Soft condensed matter*
17. S. Tautz: *Surfaces of condensed matter*
18. H. Bruus: *Theoretical microfluidics*
19. C.L. Dennis, J.F. Gregg: *The art of spintronics: an introduction*
21. T.T. Heikkilä: *The physics of nanoelectronics: transport and fluctuation phenomena at low temperatures*
22. M. Geoghegan, G. Hadziioannou: *Polymer electronics*

ATOMIC, OPTICAL, AND LASER PHYSICS

7. C.J. Foot: *Atomic physics*
8. G.A. Brooker: *Modern classical optics*
9. S.M. Hooker, C.E. Webb: *Laser physics*
15. A.M. Fox: *Quantum optics: an introduction*
16. S.M. Barnett: *Quantum information*
23. P. Blood: *Quantum confined laser devices*

PARTICLE PHYSICS, ASTROPHYSICS, AND COSMOLOGY

10. D.H. Perkins: *Particle astrophysics, second edition*
11. Ta-Pei Cheng: *Relativity, gravitation and cosmology, second edition*
24. G. Barr, R. Devenish, R. Walczak & T. Weidberg, *Particle physics in the LHC era*

STATISTICAL, COMPUTATIONAL, AND THEORETICAL PHYSICS

12. M. Maggiore: *A modern introduction to quantum field theory*
13. W. Krauth: *Statistical mechanics: algorithms and computations*
14. J.P. Sethna: *Statistical mechanics: entropy, order parameters, and complexity*
20. S.N. Dorogovtsev: *Lectures on complex networks*

Particle Physics in the LHC Era

G. Barr, R. Devenish, R. Walczak, & T. Weidberg

Department of Physics, University of Oxford

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Giles Barr, Robin Devenish, Roman Walczak, & Tony Weidberg 2016

The moral rights of the authors have been asserted

First Edition published in 2016

Impression: 1

It is permitted to copy, distribute, transmit, and adapt this work, including for commercial
purposes, provided that appropriate credit is given to the creator and copyright holder, a link is
provided to the licence, and any changes made to the work are properly indicated. You may do so
in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



This is an open access publication. Except where otherwise noted, this work is distributed under and subject to the terms of
a Creative Commons Attribution 4.0 International License (CC BY 4.0)
a copy of which is available at <https://creativecommons.org/licenses/by/4.0/>.

Enquiries concerning reproduction outside the scope of this licence
should be sent to the Rights Department, Oxford University Press, at the address above

This title has been made available on an Open Access basis
through the sponsorship of SCOAP³.



Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2015957415

ISBN 978-0-19-874855-7

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Preface

This book is designed to give a clear introduction to particle physics at a level that will be accessible to advanced undergraduate students. Most of the book concentrates on the ‘Standard Model of Particle Physics’—what it is and the experimental evidence that supports it. The book fills a gap between the more qualitative introductory texts and the more mathematically advanced graduate textbooks. Our aim is to teach the maximum amount of physics, with the minimum level of maths. This book is in the spirit of Perkins’ classic textbook but updated to the LHC era. Particle physics is an experimental science; accelerators and detectors have been essential for progress in this field. The unique feature of this book is that it gives a serious explanation of the practical side of the subject at an accessible level for undergraduates. This will provide students with some real understanding of these subjects and equip them to appreciate the many excellent graduate-level textbooks in these fields and to follow published papers. The core of the book covers the theory and experiments that underpin the Standard Model. Neutrino oscillations and flavour oscillations of the neutral strange, charm, and beauty states are explained carefully, including the violation of CP symmetry. The book covers the discovery of the Higgs at the LHC, explaining the critical issues and how one can extract such a small signal from a large background. We discuss the problems with the Standard Model that give a very strong indication that there should be physics at the TeV scale, Beyond the Standard Model (BSM). We summarize some possible BSM theories that could solve these problems and give examples of LHC searches for BSM physics. We review the evidence for dark matter and consider how LHC and other experiments are searching for it, and we look at the evidence for dark energy and its theoretical consequences.

Each chapter has questions to help students deepen their understanding of the subject, some of which are based on those used in teaching this subject at Oxford as a 4th-year physics major option course.

The official OUP website for this book is <http://ukcatalogue.oup.com/product/9780198748557.do>. This contains a link to a website ppLHCera.physics.ox.ac.uk maintained by the authors. We are maintaining a list of errata on this website and we would appreciate receiving corrections via the link on the OUP website. Many new results will be appearing over the next few years. We provide links to Particle Data Group reviews and to websites for some of the current experiments. Finally, suggested solutions are available to course instructors: a request form is available on the OUP book website.

Acknowledgements

Many people have contributed significantly to this book. First, John Cobb and Malcolm John wrote preliminary drafts of Chapters 10 and 11 on flavour and neutrino oscillations and Chapter 2 on mathematical methods. Jim Libby (now at the Indian Institute of Technology, Madras) created many of the stylish plots for the quark model in Chapter 5. Several colleagues at Oxford read individual chapters or sections and provided very useful corrections and suggestions: Alan Barr, Chris Hays, Todd Huffman, Cigdem Issever, Ivan Konoplev, Andrei Starinets, Georg Viehhauser, and Alfons Weber. Amanda Cooper-Sarkar provided invaluable advice on parton distribution functions and the data to produce Fig. 9.13. Steve Biller and Armin Reichold gave useful suggestions on the subject of neutrino oscillations. Hugo Beauchemin (Tufts) provided very useful feedback on Chapter 4 on detectors. Many Oxford students provided vital corrections and useful suggestions. An incomplete list is given here, with apologies to those we have forgotten: Callum Brodie, Rehan Deen, Patrick Dunne, Scott Melville, Jack Miller, Ricky Nathvani, Jack Sennett, Jakub Sikorowski, Jacob Smith, Elliot Reynolds, Jack Weston, and Andrew Yeomans. Members of our families helped as well. Anna Walczak made Figs. 3.3 and 3.7, and Przemysław Walczak helped make figures for Chapters 3, 4, and 6. Sheila McKinnon helped with proof reading.

We are indebted to Sonke Adlung of OUP for his continuous support and encouragement over the many years this book has taken to complete and we thank Harriet Konishi of OUP for guiding us through the final stages of production of the book. We thank Mac Clarke for the very thorough and excellent copyediting of the text. We thank Kaarkuzhali Gunasekaran of Integra-PDY for very careful work in typesetting.

Needless to say, all residual mistakes are the responsibility of the current authors. We would be very grateful to receive any corrections and comments from readers. We will keep an up-to-date list of errata on the book's website.

Figure permissions

We are grateful to the following publishers and organizations, who retain the original copyright, for permission to include the following figures (our figure numbers).

Addison-Wesley. Fig. 9.15 is reproduced from *Introduction to High Energy Physics* (3rd edn), D. H. Perkins, 1987.

The American Physical Society. Fig. 9.3 is reproduced from *Physical Review*. Figs. 2.4, 8.22, 8.23, and 11.8. is reproduced from *Physical Review D*. Figs. 5.9, 5.10, 8.26, 9.18, 9.33, 11.4, and 11.10 is reproduced from *Physical Review Letters*.

Cambridge University Press. Fig. 9.12 is reproduced from *Neutrino Physics* (2nd edn), K. Winter, 2000.

Elsevier. Figs. 4.8, 4.27, 8.5, and 13.29 is reproduced from *Nuclear Instruments and Methods*. Figs. 6.1 and 9.22 is reproduced from *Physics Letters*. Figs. 2.2, 2.6, 9.11, 10.7, 12.10, 12.12, 12.13, 12.15, 12.16, and 12.18 is reproduced from *Physics Letters B*. Figs. 8.10, 8.15, 8.18, 8.28, and 9.2 is reproduced from *Physics Reports*.

Institute of Physics Publishing. Figs. 9.28 and 9.29 is reproduced from *Reports on Progress in Physics*.

The KTeV Collaboration: Fig. 10.8.

World Scientific. Fig. 9.1 is reproduced from *Nuclear Structure I*, A. Bohr and B. Mottelson, 1998.

Other figures are covered by publishers' open access agreements or are the authors' own. For those not the authors' own, the reference to the individual publication is given in the caption to the figure.

Acronyms

ATLAS	A toroidal LHC apparatus
BSM	Beyond the standard model
BR	Branching ratio
CC	Charged current (i.e. W^\pm -mediated weak interactions)
CERN	European Particle Physics Laboratory, Geneva, Switzerland
CKM	Cabibbo–Kobayashi–Maskawa, quark mixing matrix
CMS	Centre-of-mass system
CMS	Compact muon spectrometer, collider detector at the LHC
DESY	Deutsches Elektronen Synchrotron, particle physics laboratory in Hamburg, Germany
DIS	Deep inelastic scattering
ES	Elastic scattering
EM	Electromagnetism
FNAL	Fermi National Accelerator Laboratory, Batavia, Illinois, USA
HERA	Hadronen-Elektronen-Ring-Anlage (hadron–electron collider at DESY)
GIM	Glashow–Iliopoulos–Maiani, necessity for the charm quark
ISR	Initial-state radiation
H1	Collider detector at HERA
LEP	Large Electron–Positron Collider at CERN
LHC	Large Hadron Collider at CERN
LHCb	LHC beauty experiment
MINOS	Main injector oscillation experiment at FNAL
MSW	Mikheyev–Smirnov–Wolfenstein, mass effect in neutrino oscillations
NC	Neutral current (i.e. γ^* - or Z^0 -mediated interactions)
NR	Non-relativistic
PEP, PEP-II	Positron–electron colliders at SLAC
PETRA	Positronen-Elektronen-Tandem-Ring-Anlage, electron–positron collider at DESY
PDF	Parton (momentum) density function

PNMS	Pontecorvo–Maki–Nakagawa–Sakata, neutrino mixing matrix
pQCD	Perturbative quantum chromodynamics
Q^2	Negative of 4-momentum transfer squared in deep inelastic scattering
QCD	Quantum chromodynamics
QED	Quantum electrodynamics
QPM	Quark–parton model
RICH	Ring imaging Čerenkov radiation detector
RQF	Relativistic quantum fields
RQM	Relativistic quantum mechanics
SLAC	Stanford Linear Accelerator Center, Stanford, California, USA
SM	Standard Model (of particle physics)
SNO	Solar neutrino oscillation experiment, Sudbury, Ontario, Canada
TASSO	Collider detector at PETRA (DESY)
Tevatron	Proton–Antiproton collider at FNAL
T2K	Tokai to Kamioka, neutrino oscillation experiment
ZEUS	Collider detector at HERA (DESY)

Contents

1	Introduction	1
1.1	Units	2
1.2	Early days	2
1.2.1	Particles and forces	3
1.2.2	Group theory in particle physics	4
1.2.3	Particles	5
1.2.4	Forces	6
1.3	Diagrams	7
1.4	Accelerators, colliders, and detectors	8
1.4.1	Accelerators	8
1.4.2	Colliders	10
1.5	Detectors	10
1.6	Open questions	11
1.7	Chapter outline	12
1.8	How to read this book	13
	Chapter summary	13
	Further reading	14
2	Mathematical methods	15
2.1	Discrete symmetries	15
2.1.1	Spatial parity	15
2.1.2	Charge conjugation	16
2.1.3	Time reversal	17
2.1.4	J^{PC} of hadrons	18
2.1.5	Useful examples	18
2.2	Addition of angular momentum	19
2.2.1	Angular momentum in quantum mechanics	19
2.2.2	Addition and Clebsch–Gordan coefficients	21
2.2.3	Calculation of Clebsch–Gordan coefficients	21
2.3	Spatial rotations	23
2.4	Lorentz invariance	29
2.4.1	Invariant variables	30
2.4.2	Rapidity	30
2.5	Transitions and observables	31
2.5.1	Phase space and decay rates	32
2.5.2	Decay rate	34
2.5.3	Cross section	39
2.5.4	Breit–Wigner	40

2.6	Luminosity and event rates	42
2.7	Group theory	43
2.7.1	Lie groups	44
2.7.2	$U(n)$ and $SU(n)$	44
2.7.3	$SU(2)$	45
2.7.4	Combining states	46
	Chapter summary	46
	Further reading	46
	Exercises	47
3	Accelerators	48
3.1	Radiofrequency acceleration	48
3.1.1	Electric and magnetic fields	48
3.1.2	Circular accelerators and synchronicity	49
3.1.3	Accelerating-cavity design and Q factor	51
3.1.4	Synchrotron radiation energy loss	53
3.1.5	Linear accelerators	53
3.1.6	Klystrons	54
3.2	Beam optics	54
3.2.1	Magnetic lenses	54
3.2.2	Beam trajectories and phase space	57
3.3	LHC dipole magnets	59
3.3.1	Engineering design details	61
3.4	Colliders and fixed-target accelerators	62
3.5	Luminosity	63
3.6	$\bar{p}p$ colliders	65
3.6.1	CERN $\bar{p}p$ collider	66
3.6.2	Tevatron $\bar{p}p$ collider	67
3.7	Future accelerators	67
	Chapter summary	68
	Further reading	68
	Exercises	69
4	Particle detectors	71
4.1	Introduction	71
4.2	Overview of collider detectors	72
4.3	Particle interactions with matter	74
4.3.1	Ionization	74
4.3.2	Multiple scattering	76
4.3.3	Electromagnetic interactions	77
4.3.4	Čerenkov radiation	79
4.3.5	Transition radiation	79
4.3.6	Hadronic interactions	79
4.4	Signal generation	81
4.4.1	Moving charges	81
4.4.2	Scintillators	84
4.5	Photon detection	85
4.6	Detectors for charged-particle tracks	86

4.6.1	Wire chambers	87
4.6.2	Silicon detectors	92
4.6.3	Tracker performance	97
4.7	Detectors for particle jets	98
4.7.1	Electromagnetic calorimeter	98
4.7.2	Homogeneous calorimeters	99
4.7.3	Sandwich calorimeters	100
4.7.4	Resolution	102
4.7.5	Hadronic calorimeter	103
4.8	Detectors for particle identification	105
4.8.1	Particle identification with Čerenkov detection	105
4.8.2	Particle identification with transition radiation	105
4.8.3	Particle identification with ionization	106
4.9	Magnetic fields	106
4.9.1	Magnetic fields for trackers	106
4.9.2	Magnetic fields for muon spectrometers	107
4.10	Trigger	108
4.10.1	LHC triggers	109
4.11	Examples of detector systems	109
4.11.1	Collider detectors	109
4.11.2	Neutrino detectors	111
	Chapter summary	112
	Further reading	112
	Exercises	112
5	Static quark model	115
5.1	Spin $\frac{1}{2}$	115
5.1.1	Combining two spin- $\frac{1}{2}$ particles	116
5.1.2	Combining three spin- $\frac{1}{2}$ particles	117
5.2	The quark model of hadrons	118
5.2.1	Isospin	119
5.2.2	Strangeness and expansion to SU(3)	122
5.2.3	Mesons	123
5.2.4	Baryons	125
5.2.5	Deriving the complete spin-flavour wavefunction	127
5.3	Heavy quarks	128
5.3.1	The charm quark	129
5.3.2	The beauty quark	130
5.3.3	The top quark	130
5.3.4	Charm and beauty states	130
5.3.5	Heavy $Q\bar{Q}$ systems	132
5.3.6	Charmonium	132
5.3.7	Comparison with positronium	134
5.3.8	Bottomonium	136
5.4	Exotic hadrons	137
	Chapter summary	138
	Further reading	138
	Exercises	139

6	Relativistic quantum mechanics	140
6.1	Special relativity	140
6.1.1	Spinors	142
6.2	One-particle states	147
6.2.1	Fields and probability amplitudes	148
6.3	The Klein–Gordon equation	149
6.3.1	The Feynman–Stueckelberg interpretation of negative-energy states	151
6.4	The Dirac equation	155
6.4.1	Free-particle solutions	159
6.4.2	Chirality \neq helicity	161
6.4.3	Helicity conservation and interactions via currents	163
6.4.4	P , T , and a comment on C	166
6.4.5	Electromagnetic interactions and the non-relativistic limit	169
6.5	Gauge symmetry	172
6.5.1	Covariant derivative	172
6.5.2	Gauge invariance in electromagnetism	174
6.5.3	The Aharonov–Bohm effect	174
6.5.4	Interactions from gauge symmetry	175
	Chapter summary	178
	Further reading	178
	Exercises	179
7	Weak interactions	181
7.1	Fermi theory	182
7.2	Weak interactions of leptons	183
7.2.1	Lepton number	183
7.2.2	Feynman rules	183
7.2.3	Universality	185
7.2.4	V–A	186
7.2.5	Parity violation	187
7.2.6	Currents and fields	187
7.3	Weak interactions including quarks	189
7.3.1	Cabibbo theory	189
7.3.2	GIM mechanism, flavour-changing neutral currents	189
7.3.3	CKM matrix	190
7.3.4	Decays of hadrons containing heavy quarks	192
7.4	Introduction to electroweak unification	193
7.4.1	Electroweak unification procedure	193
7.4.2	Weak neutral currents	199
7.4.3	Masses of W and Z bosons	200
7.4.4	The standard model, how good is it?	201
	Chapter summary	202
	Further reading	203
	Exercises	203

8	Experimental tests of electroweak theory	205
8.1	Neutrinos	205
8.2	Charged currents	206
8.2.1	Measurements of CKM matrix elements	207
8.3	Neutral currents	208
8.4	Physics at e^+e^- colliders	208
8.4.1	Detailed look at the detectors	209
8.4.2	Aspects of a physics analysis	213
8.4.3	Monte Carlo simulation	213
8.4.4	Physics at LEP	215
8.4.5	The Z line shape	216
8.4.6	Z mass	217
8.4.7	The Z width; number of neutrinos	217
8.4.8	LEP beam energy measurement	218
8.4.9	Cross sections and forward–backward asymmetries at the Z	219
8.4.10	LEP luminosity measurement	221
8.4.11	Measurements at LEP2, above $\sqrt{s} = M_Z$	221
8.4.12	W^+W^- production	222
8.4.13	$\sigma(e^+e^- \rightarrow W^+W^-)$	223
8.5	W and Z physics at hadron colliders	224
8.5.1	W and Z discovery	224
8.5.2	W mass determination at the Tevatron	225
8.5.3	Width of the W	228
8.6	Top-quark physics	229
8.6.1	Top-quark discovery	229
8.6.2	Top-quark mass measurement	230
8.6.3	Top-quark production cross sections	231
8.7	Summary	232
	Chapter summary	233
	Further reading	233
	Exercises	233
9	Dynamic quarks	235
9.1	Rutherford scattering	235
9.2	Scattering from nucleons	239
9.3	Quark–parton model	240
9.3.1	Kinematics of deep inelastic scattering	240
9.4	Neutrino interactions	242
9.4.1	Cross section for neutrino–electron elastic scattering	243
9.4.2	Neutrino–quark scattering	245
9.4.3	Neutrino–nucleon cross sections	246
9.4.4	Parton distribution functions	247
9.5	Charged-lepton probes	251
9.5.1	Electron–muon elastic scattering	252
9.5.2	Electron–quark elastic scattering	254
9.5.3	Electron–nucleon deep inelastic scattering	254

9.5.4	Further tests of the QPM	256
9.5.5	Electroweak unification at HERA	258
9.6	QCD introduction	259
9.6.1	Direct evidence for gluons	260
9.6.2	Number of colours	261
9.6.3	QCD	262
9.6.4	Running coupling constants	266
9.6.5	Experimental tests of the gauge structure of QCD	269
9.6.6	Experimental fits to the quark distribution functions	271
9.6.7	The gluon distribution function	271
9.7	Hadron–hadron collisions	272
9.7.1	Drell–Yan	273
	Chapter summary	274
	Further reading	275
	Exercises	275
10	Oscillations and CP violation in meson systems	277
10.1	Symmetries	278
10.2	Neutral kaon decays and K_1 and K_2	278
10.3	Mass differences of neutral mesons	281
10.4	Flavour oscillations	282
10.4.1	K^0 – \bar{K}^0 oscillations	283
10.4.2	D^0 – \bar{D}^0 oscillations	284
10.4.3	B^0 – \bar{B}^0 mixing and oscillations	285
10.4.4	B_s oscillations	287
10.4.5	Regeneration	288
10.5	CP violation (part 1)	289
10.5.1	Discovery of CP violation	289
10.5.2	Semileptonic charge asymmetry	291
10.5.3	CP violation in K^0 decay	292
10.6	CP violation in the Standard Model	294
10.6.1	Mixing with CP violation	296
10.7	CP violation (part 2)	297
10.7.1	CP violation in time-dependent asymmetries	297
10.7.2	B-factories	299
10.7.3	LHCb detector	301
10.8	LHCb measurements	302
	Chapter summary	304
	Further reading	304
	Exercises	304
11	Neutrino oscillations	306
11.1	Introduction	306
11.1.1	Neutrino masses	306
11.2	Neutrino states	307
11.3	Two-flavour oscillations	308
11.4	Evidence for neutrino oscillations	310

11.4.1	Atmospheric neutrinos	311
11.4.2	Laboratory confirmation of atmospheric neutrino oscillation	314
11.4.3	Solar neutrinos	315
11.4.4	MSW effect	320
11.4.5	Confirmation of solar neutrino oscillations	322
11.5	Three (or more)-flavour oscillations	323
11.5.1	Generalized oscillation probabilities	323
11.5.2	Three-flavour oscillations	326
11.5.3	Measurement of the mixing angle θ_{13}	328
11.5.4	Matter–antimatter asymmetry	329
	Chapter summary	331
	Further reading	331
	Exercises	331
12	The Higgs boson	333
12.1	Local gauge invariance	334
12.2	Spontaneous symmetry breaking	334
12.3	Higgs mechanism—the simplified story	336
12.4	Lagrangians	337
12.4.1	Lagrangians in classical mechanics	337
12.4.2	Lagrangians in quantum mechanics	339
12.5	Higgs mechanism—more mathematical	339
12.6	Higgs discovery	342
12.6.1	$\gamma\gamma$ channel	344
12.6.2	ZZ^* channel	345
12.6.3	WW^* channel	346
12.6.4	Statistical significance	347
12.7	Coupling to fermions	349
12.8	Determination of the spin and parity of the new boson	351
12.9	Outlook	353
	Chapter summary	353
	Further reading	354
	Exercises	354
13	LHC and BSM	356
13.1	LHC and Standard Model physics	356
13.2	LHC triggers	358
13.3	SM measurements at the LHC	359
13.3.1	Top-quark production	365
13.4	Beyond the Standard Model physics	367
13.4.1	Supersymmetry	368
13.4.2	R -parity	371
13.4.3	Other BSM theories	372
13.5	Experimental searches for BSM physics at the LHC	373
13.5.1	Jet production	373
13.5.2	Lepton pair resonances	373
13.5.3	SUSY searches	373

13.5.4 Summary of searches for new physics	379
13.6 Linear collider	379
13.7 Dark matter	381
13.7.1 Astrophysical evidence for dark matter	382
13.7.2 Direct dark matter detection	383
13.7.3 Dark matter annihilation	385
13.8 Dark energy	385
13.8.1 Theoretical implications	386
Chapter summary	387
Further reading	388
Exercises	388
References	391
Index	399

Introduction

1

The aim of this book is to provide a practical introduction to particle physics in the LHC era at the level of an advanced undergraduate or introductory graduate course. It fills a gap between qualitative introductory texts and advanced texts based on relativistic quantum field theory. We give a clear and concise explanation of key theoretical concepts and their grounding in experimental measurements, with as little use of advanced mathematical techniques as possible. The exceptions are a fairly detailed coverage of exact and broken symmetries and gauge invariance. The language and techniques of relativistic quantum field theory are not used, but relativistic quantum mechanics is covered, focusing on the Klein–Gordon and Dirac equations. The book focuses on the physics of colliders, particularly those delivering the highest energies: proton–proton (LHC), electron–positron (LEP and ILC), and electron–proton (HERA). Experiments and results from older and/or lower-energy electron–positron colliders are discussed when necessary, for example the so-called B-factories. Fixed-target experiments, particularly those using neutrino beams and those studying neutral meson oscillation phenomena (K^0 , B^0) are outlined. Finally non-accelerator-based particle physics topics, such as the observation and measurement of solar neutrinos and the experimental search for dark matter, are described briefly.

This chapter introduces the fundamental particles and the forces with which they interact. We will use ‘natural units’ and these are defined in the next section. Accelerators and colliders are introduced in a historical context that makes clear how much of our current understanding of fundamental particles and forces relies on the steady increase in interaction energy made possible by advances in accelerator technology. An important feature of this book is a description of how modern electronic detectors work—particularly the large detectors built to study proton–proton collisions at the multi-TeV scale of the LHC. Coupled with this is the enormous increase in computing power over the last half-century and the organization of this power on a global scale using the worldwide web,¹ which enables scattering events to be selected and reconstructed in almost real time.

The chapter ends with a brief resume of the rest of the book.

1.1 Units	2
1.2 Early days	2
1.3 Diagrams	7
1.4 Accelerators, colliders, and detectors	8
1.5 Detectors	10
1.6 Open questions	11
1.7 Chapter outline	12
1.8 How to read this book	13
Chapter summary	13
Further reading	14

¹This is known as GRID computing. A local example used by the authors is GRIDPP: <http://www.gridpp.ac.uk/>.

1.1 Units

Studying matter at subnuclear scales requires interactions at very high energies. SI units are very useful in many contexts, but for this subject they would require us to keep track of quantities, particularly energies, to large negative powers. Particle physics uses the **natural units** of MeV or GeV ($\equiv 1000$ MeV) for energy, the femtometre (also known as the fermi, $1 \text{ fm} = 10^{-15} \text{ m}$) for distance, and the barn ($1 \text{ b} = 10^{-28} \text{ m}^2$) for area (used for cross sections).²

²It is said that the name ‘barn’ originated when early measurements of nuclear scattering cross sections were larger than expected: ‘as big as barn doors’.

By using natural units, we can set $\hbar = c = 1$ in our calculations. So, for example, the familiar relation between energy and momentum in special relativity becomes simply $E^2 = p^2 + m^2$. In natural units, mass, energy, and momentum have the same units, which simplifies dimensional analysis checking. At the end of a calculation, we might need to convert the answer to practical units, which we can do very simply by using the conversion factors $\hbar c = 197.3 \text{ MeV fm}$ and $(\hbar c)^2 = 0.3894 \text{ GeV}^2 \text{ mb}$.

We will always use natural units in this book, unless explicitly indicated otherwise.

1.2 Early days

Particle physics had its roots in nuclear physics and cosmic-ray physics. Its remit is the study of the fundamental building blocks of matter and the interactions between them, particularly the strong, electromagnetic, and weak interactions. The gravitational interaction does not play a significant role in most of the topics covered in this book.³

³Compare gravitational and electrostatic forces between two protons at a separation of 10^{-12} m .

In the 1930s, it appeared that atoms and nuclei could be understood in terms of a rather small number of constituents: the proton and neutron; the electron and positron; the neutrino and antineutrino. Electromagnetic interactions were assumed to remain described by the theories of Faraday and Maxwell at subatomic distance scales, but consistent relativistic calculations required the development of quantum field theory and, in particular, an understanding of renormalization (see Section 1.2.1). Weak interactions were more of a problem. The contact interaction developed by Fermi was very successful in bringing order to a wide range of phenomena, but it was not renormalizable and hence did not allow reliable higher-order calculations. The strong nuclear force has a range of a few fermi (10^{-15} m), the scale of a nucleus. The strength of the electromagnetic interaction is given by the fine-structure constant, $\alpha = e^2/4\pi\hbar c$, with a value $\frac{1}{137}$. Defining a ‘strong nuclear charge’ α_S similarly gives $\alpha_S \sim 1$ at a distance of a few fermi. The scale of weak interactions is given by the the Fermi constant $G_F/(\hbar c)^3 = 1.166 \times 10^{-5} \text{ GeV}^{-2}$.

This simple picture could not accommodate the discovery of new particles by cosmic-ray physicists: the muon by Anderson and Neddermeyer in 1936 using a cloud chamber within a magnetic field; the pion in 1947 by Powell’s group in Bristol using specially developed photographic emulsion; and ‘strange particles’ with V-shaped or kinked tracks by

Rochester and Butler (1946–47) using a coincidence-counter-controlled cloud chamber. What was strange about the new particles was a much longer lifetime than would be expected for a ‘normal’ strongly interacting particle with a comparable mass. A new quantum number, strangeness, was introduced independently by Gell-Mann and by Nishijima in 1953 to explain this. Their proposal was that strangeness was conserved in strong and electromagnetic interactions but not in weak interactions. The lightest strange particles could only decay by a strangeness-changing weak interaction.

This was just the beginning. In the early 1950s, the development in the USA and Europe of synchrotrons capable of delivering particle beams with GeV energies and of the bubble chamber led to a plethora of new short-lived particles.

1.2.1 Particles and forces

The forces that we are most familiar with at a macroscopic level are gravitation and electromagnetism, and at this level a particle may be defined roughly as a ‘point-like’ object that has a well-defined mass and charge. This also works reasonably well at the scale of atoms (10^{-8} m), at which electrons, protons, and neutrons can be considered point-like. In quantum mechanics and quantum field theory, a force is described by the exchange of a field-quantum—in the case of electromagnetism, the photon (γ). One of the consequences of Dirac’s attempts to find a physical explanation for the troublesome negative-energy solutions of his otherwise very successful equation describing the electron was the prediction of *antiparticles* (1930–31), in particular the positron. An antiparticle is a particle with the same mass and spin as a particle, but with opposite electric charge. A neutral particle can be its own antiparticle, for example the photon. Feynman⁴ has given a simple but elegant argument that antiparticles are a necessary outcome of a relativistically invariant description of particle interactions. Shortly after Dirac’s prediction, the positron was discovered by Anderson in 1932. By the late 1940s, the development of *renormalized* quantum field theories⁵ gave a consistent way to handle the infinities that seemed inherent in any quantum description of particle creation and annihilation. The paradigm is quantum electrodynamics (QED), for which the mass and electric charge of the electron are two of the parameters that are renormalized, the so-called vacuum energy⁶ being the third. After renormalization, QED provides a theory that has enabled amazingly precise calculations of quantities such as the magnetic moment of the electron.⁷

At a deep level, much of the thrust of particle physics in the twentieth century was to find out if the strong and weak nuclear forces could be described by renormalized quantum fields, and, if so, to discover the related field quanta. We now know that this is the case, with the W^\pm and Z^0 providing the weak force and the eight massless gluons (g) the strong force. To describe these interactions, more complicated field theories are required, but they have been shown to be renormalizable.

⁴‘The reason for antiparticles’, in the 1986 Dirac Memorial lectures—see the further reading at the end of this chapter.

⁵A renormalizable theory is one in which infinities to all orders of perturbation theory can be absorbed by the redefinition of a finite number of the parameters of the theory, such as masses and coupling constants. These parameters are then fixed from experimental observation.

⁶A consequence of the quantum time–energy uncertainty relation, vacuum energy ΔE can exist for a time Δt , provided $\Delta E \Delta t < \hbar$.

⁷Quantum field theory is beyond the remit of this book, but some introductory texts are listed in the further reading at the end of Chapter 6.

1.2.2 Group theory in particle physics

Group theory is the mathematical description of patterns, both of physical structures such as crystals and more abstractly of groups of related objects, for example particles with similar properties such as the pions (π^\pm, π^0). Group theory also plays an essential role in the mathematical structure describing the forces and interactions of fundamental particles.⁸ The great benefit of group theory is that it provides much of the mathematical apparatus needed to exploit the underlying symmetries of the fundamental forces.

⁸A more detailed account of the group theory that we need is given at the end of Chapter 2.

⁹The subscript ‘flavour’ is to distinguish this symmetry from the exact $SU(3)_{\text{colour}}$ of QCD. The subscripts may be omitted if the context is unambiguous.

¹⁰Quarks, antiquarks, and gluons.

As we shall explain in detail in Chapter 5, the mesons and baryons, composed of the u , d , and s quarks, occur in patterns of a symmetry known as⁹ $SU(3)_{\text{flavour}}$. This symmetry is approximate because of the difference in masses of the three quarks. It contains an $SU(2)$ subgroup known as isospin composed of hadrons with only u and d quarks. The pion states just mentioned form an isospin triplet and the neutron and proton an isospin doublet.

The strong interaction among the constituents¹⁰ of hadrons is based on an exact $SU(3)$ group structure with three ‘colour’ charges. This is the theory of quantum chromodynamics (QCD), which underpins the physics described in Chapter 9. The space–time structure of QCD is similar to that of QED, with an inverse-square-law force. Like the photon, the gluons are massless particles, but, unlike the photon, the gluons carry a colour charge—there are eight gluons. Further explanation is given in Chapter 9.

Group theory in electroweak unification

The most complicated use of group theory is in describing the ‘unification’ of the weak and electromagnetic interactions to form the electroweak theory of the Standard Model. It is complicated because QED is a spatial parity-conserving force whereas the weak force does not respect this symmetry. Electrodynamics has another important feature—it is ‘gauge-invariant’. Maxwell’s equations are unchanged by a change in the electromagnetic 4-vector potential $A_\mu \rightarrow A_\mu - \partial_\mu \Lambda$, where Λ is a scalar function. Under a gauge transformation, a wavefunction changes by a phase $\psi \rightarrow \psi e^{-ie\Lambda}$. In the language of group theory, this is a $U(1)$ symmetry. This is much more than just a mathematical curiosity, since the replacement of 4-momentum of a charged particle $p_\mu \rightarrow p_\mu - eA_\mu$ in the (classical) equations of motion generates the correct form for the electromagnetic interaction.¹¹

¹¹This is covered in the discussion of gauge symmetry in Chapter 6.

¹²For an ultrarelativistic particle, ‘left-handed’ refers to the component of its spin projected along a direction opposite to that of its 3-momentum.

¹³The subscript L is for ‘left-handed’, but also indicates that this is not the approximate $SU(2)$ of nuclear isospin mentioned above.

The quanta of the weak force are the charged W^\pm spin-1 bosons interacting via their ‘left-handed’¹² states and the Z^0 , which couples to both left- and right-handed states but with different strengths. As both the photon and Z^0 are spin-1 states with zero electric charge, they can interfere—with a strength given by a mixing angle θ_W (the ‘weak mixing angle’ or ‘Weinberg angle’). The group structure of the left-handed states is that of¹³ $SU(2)_L$. Using group-theoretical language, this synthesis of electrodynamics with the weak interaction is based on a $U(1) \otimes SU(2)_L$

group structure. In an analogous, but more complicated, procedure to that described above for QED, the full electroweak interaction can be generated by a suitable gauge transformation. Electroweak unification is covered in Section 7.4.

1.2.3 Particles

Particles, including the force quanta, are classified according to their spin and interactions. *Leptons* are spin- $\frac{1}{2}$ fermions that do not interact via the strong force: the electron (e) and associated neutrino (ν_e) provide the paradigm. Two further sets or *generations* have been discovered: the muon (μ) and muon-neutrino (ν_μ) and the tau (τ) and tau-neutrino (ν_τ). To account for the non-observation of decay modes such as $\mu \rightarrow e\gamma$ and $\tau \rightarrow \mu\gamma$, each generation of lepton pairs is given a lepton number, L_ℓ . For each lepton, there is a corresponding antiparticle with opposite signs for charge Q and lepton number. The properties of the leptons are summarized in Table 1.1.

As will be explained in Chapter 8, the number of neutrino species, N_ν , is given by the width of the Z^0 vector boson and is consistent with a value of 3.

Strongly interacting particles (*hadrons*) are composed of quarks and antiquarks bound tightly in $q\bar{q}$ (*meson*) or qqq (*baryon*) combinations by the colour field of QCD. Free quarks have never been observed directly, although there is evidence that they may become unbound in a quark–gluon plasma, which is being studied using heavy-ion collisions. Quarks carry fractional electric charge, $\frac{2}{3}e$ or $-\frac{1}{3}e$, where e is the charge of the positron. There are six quarks, grouped in charge ($\frac{2}{3}, -\frac{1}{3}$) pairs: (d, u), (s, c), (b, t). The (d, u) pair form a strong isospin doublet. The details are given in Table 1.2. All quarks have $J^P = \frac{1}{2}^+$. It is worth noting that the three pairs of quarks (d, u), (s, c), (b, t) of increasing mass are matched by the three lepton pairs (e, ν_e), (μ, ν_μ), (τ, ν_τ).

State	Q	Mass	L_e	L_μ	L_τ	Lifetime
e^-	-1	0.511 MeV	+1	0	0	$> 4.6 \times 10^{26}$ years
ν_e	0	< 2 eV	+1	0	0	Stable
μ^-	-1	105.7 MeV	0	+1	0	$2.197034(21) \times 10^{-6}$ s
ν_μ	0	< 0.19 MeV	0	+1	0	Stable
τ^-	-1	1776.82 ± 0.16 MeV	0	0	+1	$(290.6 \pm 1.0) \times 10^{-15}$ s
ν_τ	0	< 18.2 MeV	0	0	+1	Stable

Table 1.1 Lepton properties.

	Q	Mass	I	I_z
d	$-\frac{1}{3}$	4.1–5.8 MeV	$+\frac{1}{2}$	$-\frac{1}{2}$
u	$+\frac{2}{3}$	1.7–3.3 MeV	$+\frac{1}{2}$	$+\frac{1}{2}$
s	$-\frac{1}{3}$	80–130 MeV	0	0
c	$+\frac{2}{3}$	1.18–1.34 GeV	0	0
b	$-\frac{1}{3}$	$4.19^{+0.18}_{-0.06}$ GeV	0	0
t	$+\frac{2}{3}$	172.0 ± 1.6 GeV	0	0

Table 1.2 Quark properties.

1.2.4 Forces

The force carriers of the Standard Model occur in two independent sectors:

- the electroweak (EW) sector, with the photon, W^\pm and Z^0 linked by the $U(1) \otimes SU(2)$ symmetry;
- the strong quantum chromodynamic (QCD) sector, with an $SU(3)$ colour octet of massless gluons.

All force carriers in the Standard Model are spin-1 bosons and their properties are summarized in Table 1.3.

Sector	Q	Colour charge	Mass	Width	J^P	
EW	W^\pm	± 1	0	80.399(23) GeV	2.085(42) GeV	1
	Z^0	0	0	91.1876(21) GeV	2.4952(23) GeV	1
	γ	0	0	0	0 (stable)	1^-
Strong	g	0	$SU(3)_{\text{colour}}$ octet	0	0 (stable)	1^-

Table 1.3 Force carriers.

For the symmetries to be exact, the particles are assumed to be initially massless, with the Higgs mechanism being invoked to generate particle masses (for all except the photon and gluons) while preserving the underlying symmetry structure. This mechanism requires the existence of a Higgs particle and there is now strong evidence from the Large Hadron Collider at CERN for the existence of at least one Higgs boson. The details of this key discovery for completing the Standard Model are covered in Chapter 12. After a long shutdown between 2013 and 2015, the LHC will operate at the higher centre-of-mass energy of 13 TeV and with higher luminosity. Apart from studying the Higgs in greater detail, much effort will be devoted to the search for evidence of physics beyond the Standard Model.

The 3-fold colour quantum number was introduced to allow baryon wavefunctions, for example that of the Δ^{++} composed of three u quarks (spin $\frac{3}{2}$, isospin $\frac{3}{2}$), to have simultaneously the correct permutational symmetry and satisfy Fermi–Dirac statistics. QCD provides the theoretical basis for why only the ‘colourless’ qqq and $q\bar{q}$ combinations form ‘confined’ hadronic bound states. A major difference between QCD and QED is that the force carriers are ‘charged’, in that the gluons carry a colour charge. Consider a $q\bar{q}$ meson and all the possible colour combinations that a gluon exchanged between the q and \bar{q} might carry. With r, b, g denoting the $SU(3)_{\text{colour}}$ charges, from $(r, b, g) \otimes (\bar{r}, \bar{b}, \bar{g})$, one might expect nine coloured gluons. However, the three colour-neutral combinations $(r\bar{r}, b\bar{b}, g\bar{g})$ have one totally symmetric combination $\frac{1}{\sqrt{3}}(r\bar{r} + b\bar{b} + g\bar{g})$. In group-theoretical language, this corresponds to combining the 3 and $\bar{3}$ representations of $SU(3)_{\text{colour}}$: $3 \otimes \bar{3} = 8 \oplus 1$. The totally symmetric combination corresponds to the ‘1’ and would be colourless and unconfined, so it is discarded. The remaining octet of coloured gluon states are

$$r\bar{b}, \quad r\bar{g}, \quad b\bar{g}, \quad b\bar{r}, \quad g\bar{r}, \quad g\bar{b}, \quad \frac{1}{\sqrt{2}}(r\bar{r} - b\bar{b}), \quad \frac{1}{\sqrt{6}}(r\bar{r} + b\bar{b} - 2g\bar{g}).$$

Note that there are two apparently colourless states. However, these two colour states are analogous to the electrically neutral members of a strong isospin multiplet (e.g. a ρ^0)—they are not colourless.

1.3 Diagrams

Two sorts of diagrams are used in this book: Feynman diagrams and ‘quark-flow’ diagrams. Richard Feynman invented a very elegant graphical formalism that provides a considerable shortcut in calculations. Feynman rules make a direct connection between each element of a diagram and a component of the mathematical expression describing the process, derived from quantum field theory. An example of a Feynman diagram for the process $e^+d \rightarrow \bar{\nu}_e u$ via W^+ exchange is shown in Fig. 1.1. Feynman’s graphical technique was invented in the 1940s during

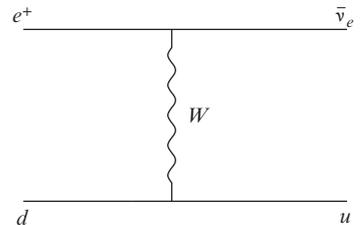


Fig. 1.1 Feynman diagram for $e^+d \rightarrow \bar{\nu}_e u$ via W^+ exchange.

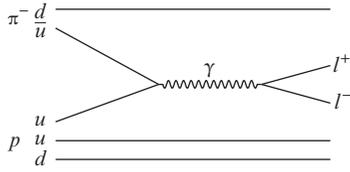


Fig. 1.2 Lepton pair production by the Drell–Yan process $\pi^- p \rightarrow n\ell^+\ell^-$.

the heroic age of relativistic quantum field theory calculations of electromagnetic interactions. More details and an outline of the ‘Feynman rules’ on how to construct a diagram are given in Section 7.2.2.

An example of a quark flow diagram is shown in Fig. 1.2. It shows the so-called Drell–Yan process $\pi^- p \rightarrow n\ell^+\ell^-$. Quark flow diagrams are not an exact calculational tool. However, they are very useful for explaining and understanding what is happening in a particle interaction. They also enable one to keep track of charges and other quantum numbers such as strangeness that may be changing but have to satisfy overall conservation laws.

1.4 Accelerators, colliders, and detectors

This section covers the essential ‘tools of the trade’ for high-energy particle physics.

Accelerators were first developed for high-energy nuclear physics in the 1930s: both ‘linear’ electrostatic devices and the first circular accelerators. After the Second World War, new technology enabled a huge increase in beam energies. A summary of accelerators and colliders in operation over the period 1950–2010 is shown in Fig. 1.3.

Detector technology has also changed a lot owing to the development of microprocessors and advanced circuit board design—some of it driven by the demands of the video-gaming industry.

1.4.1 Accelerators

The earliest particle accelerators were based on the use of a single very large potential difference to accelerate a charged particle: Van de Graaff

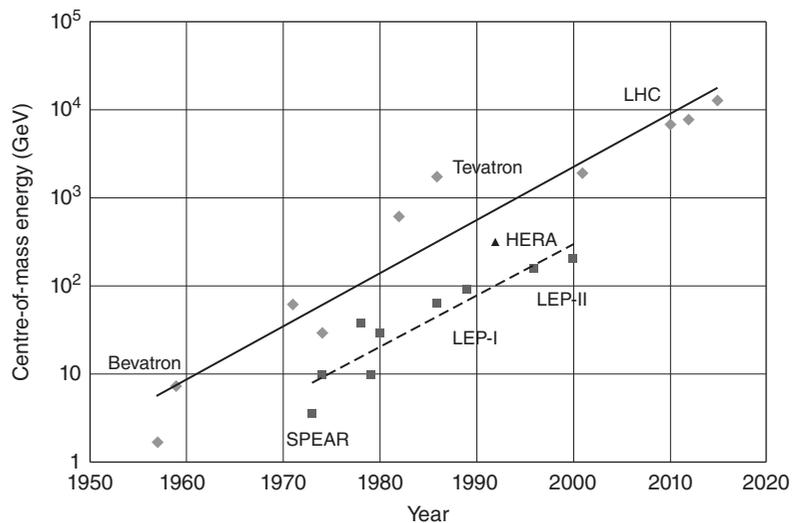


Fig. 1.3 Accelerators and colliders in use between 1950 and 2020: hadron–hadron (diamonds), electron–positron (boxes), and HERA electron–proton (triangle).

(1929) used a dielectric belt to transfer charge from a voltage source to a large spherical isolated upper terminal; Cockcroft and Walton (1937) used a series of stages to ‘multiply’ the voltage. The maximum energy was limited by electrical breakdown, typical maximum accelerating voltages being around 25 MV. Both technologies are still in use: Van de Graaff machines for research in nuclear physics and the Cockcroft–Walton multiplier as an early accelerating stage after the ion source in high-energy facilities.

It was soon realized that to get to ever higher energies, a circular device would allow the accelerating element to be used more than once. Ernest Lawrence pioneered the early development of circular accelerators in the 1930s. His machine—the cyclotron—had a circular beam with a radius that increased as it was accelerated, and the whole device was enclosed in a single large electromagnet. The largest cyclotron that Lawrence built had a diameter of 1.5 m and produced an 8 MeV proton beam.

In a modern accelerator, dipole and quadrupole magnets are used for bending and focusing, and microwave cavities for accelerating the particles. A key technical advance was the discovery of ‘phase stability’, which synchronized the accelerating voltage frequency and magnetic field strength with the rotation of the particle beam. This enabled circular machines with a beam pipe of fixed radius first to accelerate the beam (electron or proton) and then maintain it at its required operating energy. Another key advance was ‘strong focusing’, which allowed the use of much smaller vacuum pipes and hence made much larger accelerators affordable.

The first large proton accelerator at CERN¹⁴—the Proton Synchrotron (PS)—has a diameter of 200 m and a maximum beam energy of 28 GeV. Remarkably, the PS, which started operating in 1959, is still a key component of the CERN complex of accelerators. For particle physics experiments, a high-energy proton beam is extracted from the PS and then directed at a target and detector. ‘Secondary’ beams of relatively long-lived particles such as pions and kaons can also be produced from the first target and selected by more magnets and particle identification devices. Producing neutral-particle beams is a bit more challenging, since they cannot be steered by electromagnetic methods. The production of neutrino beams is discussed in Chapter 11.

Synchrotrons also accelerate electrons, such as in the original 7.5 GeV machine—the Deutsches Elektronen-Synchrotron—that gave the DESY laboratory in Hamburg its name. An extracted electron beam can be used directly for experiments or to produce a secondary photon beam by bremsstrahlung.¹⁵ Any remaining e^\pm particles can be swept from the path of the photon beam by magnets before the photon beam reaches the target.

Energy loss by synchrotron radiation from an electron beam moving in a circular orbit provides additional problems for the accelerator physicist.¹⁶ The rate of energy loss by synchrotron radiation is discussed in more detail in Chapter 3. Its effect is to limit the maximum energy at a

¹⁴The very first accelerator at CERN was the much smaller synchrocyclotron with a maximum beam energy of 600 MeV.

¹⁵Literally ‘braking radiation’, the process is $e^\pm \rightarrow e^\pm\gamma$, in which a high-energy electron emits a photon as it traverses a thin layer of matter and is deflected by the positive nuclear charge. The closely related process of ‘pair creation’, $\gamma \rightarrow e^+e^-$, will also occur; together, the two processes give rise to an electromagnetic ‘shower’ of e^\pm particles.

¹⁶Synchrotron radiation is essentially the same basic physical process as bremsstrahlung, but for much lower-energy photons (roughly X-ray energies). It is enormously useful for investigating the structure of materials, and synchrotron light sources (e.g. Diamond in the UK and the European Synchrotron Radiation Facility in France) are examples of very practical spin-offs from high-energy particle physics.

given radius. This can be countered by increasing the radius—with the ultimate result being a linear accelerator. Physicists at Stanford University first developed MeV-scale linacs for nuclear structure physics in the 1950s. Somewhat later, the Stanford Linear Accelerator Center (SLAC) was set up to build and operate a two-mile-long electron linac—the longest to date. It started operating in 1966 with a maximum beam energy of 20 GeV, which increased to nearly 50 GeV before high-energy physics experiments ceased in 1998.

1.4.2 Colliders

What matters for the study of the physics is the centre-of-mass (CMS) energy. All the above accelerators produced beams for ‘fixed-target’ experiments—as the name implies, the target is stationary. A large fraction of the beam energy has been used simply to accelerate the CMS frame in the laboratory frame of the accelerator. To exploit the maximum energy available from circular accelerators, one needs to have two counter-rotating beams and collide them head-on or nearly so. If the two beams are proton and antiproton or electron and positron, the same beam pipe can be used. An e^+e^- collider with beams of 20 GeV gives 40 GeV in the CMS frame.

The key challenge for colliders is to achieve sufficiently high luminosity to provide useful interaction rates. This difficulty has been solved as described in Chapter 3 and most of the major discoveries in particle physics in the past 40 years have been made at colliders of different types. The only exception is for physics requiring a particular type of incident particle rather than just a large interaction energy. For example, CP violation was discovered and studied in great detail using kaon beams. The most important such examples in recent years are the high-energy neutrino beams produced at CERN, Fermilab, and J-PARC in Japan.

Apart from the LHC, much of the experimental information covered in this book has come from colliders operating in the 20 years leading up to the start of data-taking at the LHC (see Fig. 1.3), in particular the Large Electron–Positron (LEP) collider at CERN (a 27 km circumference ring now containing the LHC) with CMS energies up to 209 GeV, the Tevatron proton–antiproton collider at Fermilab (4-mile circumference) with CMS energies up to 2 TeV, and the HERA electron–proton collider at DESY (11 km circumference) with 27.5 GeV electrons on 920 GeV protons providing a maximum CMS energy of 318 GeV. Older machines, particularly the e^+e^- colliders (PEP and PETRA), provided data on charm and beauty states after the 1974 ‘revolution’.¹⁷ The latter were then studied in great detail at the dedicated ‘B-factories’: KEKB in Japan and PEP-II at SLAC.

¹⁷The discovery of the J/ψ simultaneously by the Alternating Gradient Synchrotron (AGS) at Brookhaven and the SPEAR e^+e^- collider at Stanford in that year did indeed cause a revolution in our understanding of particle physics and in particular provided crucial experimental evidence in support of quantum chromodynamics as the theory of the strong interaction.

1.5 Detectors

In the era of fixed-target experiments, the bubble chamber was one of the most important types of detector. As the name implies, it exploited the

fact that boiling could be initiated in a superheated liquid by the passage of a charged particle. The liquid was kept under pressure until just before the beam arrived, at which time the chamber was expanded and then illuminated and photographed. The chamber was surrounded by a magnetic field to bend charged-particle trajectories, thereby allowing their momentum to be determined. Bubble chamber pictures still provide a very good visual aid to understanding the kinematics of high-energy particle collisions. To obtain quantitative information, it was necessary to scan the pictures manually using specialized measuring tables to digitize the tracks. Bubble chambers could only work with pulsed beams, and many were filled with liquid hydrogen, requiring very sophisticated cryogenic engineering and safety systems. An enormous bubble chamber known as Gargamelle (a 4 m long, 2 m diameter cylinder weighing 1000 tonnes) was filled with 18 tonnes of Freon (a refrigerant) for neutrino interactions. This device was designed to find evidence for ‘weak neutral currents’, which led the way to the discovery of the Z^0 .

The alternatives to a visual device like a bubble chamber are electronic detectors. Devices such as spark and drift chambers give reasonably good spatial information on charged-particle tracks and most importantly can cope with a much higher interaction and read-out rate than a bubble chamber. Drift-chamber technology provided a sophistication that made the big devices used in collider experiments almost the equivalent of an ‘electronic bubble chamber’. However, silicon detectors offer much better resolution than drift chambers and they have now become the detectors of choice for the inner detectors at LHC, although wire chambers are still required for the very large areas in muon chambers. The energies of both charged and neutral hadrons can be measured by a calorimeter—a device providing an electronic signal proportional to the energy deposited. A traditional design is the sampling calorimeter with plates of a heavy material (such as lead) separated by space for a charge-sensitive detector. The latter can be based on liquid argon or another inert element like krypton or it can be a plastic scintillator with a photomultiplier readout. Calorimeters are also crucial for providing information that allows the separation of ‘electromagnetic particles’ (electrons and photons), hadrons, and muons. These subjects are covered in more detail in Chapter 4.

1.6 Open questions

From the 1960s on, the increases in the energy of accelerators and colliders and in the sophistication of detectors, together with powerful computers to analyse the data obtained, have provided a huge number of hadronic states with well-defined mass, width (or lifetime), spin, parity, and decay modes. The information is regularly updated and published in the ‘Review of Particle Physics’ by the much respected Particle Data Group Collaboration.¹⁸ Initially, the information was summarized on a card small enough to fit into a wallet. Now even the summary (the ‘Particle Physics Booklet’) is the size of a pocket diary and the full

¹⁸Available online at <http://pdg.lbl.gov/> or the PDG UK mirror site <http://durpdg.dur.ac.uk/lbl/>.

¹⁹See for example Steven Weinberg's 'Towards the final laws of physics', in the 1986 Dirac Memorial Lectures—the details are in the further reading.

Review is a hefty journal volume of well over 1000 pages. The Standard Model gives a very good description of this wide range of experimental information on elementary particles and their weak, electromagnetic, and strong interactions covering an energy scale up to of order 1 TeV. However, it is certainly incomplete.¹⁹ Gravity is not included and there is no explanation of why there are three 'generations' of quarks and leptons. There are nearly 20 parameters (masses, coupling constants, and mixing angles) that are not given by the Standard Model but have to be determined from experimental measurements. Other big questions crowd in—one of the most glaring is the gross matter–antimatter asymmetry of the world we inhabit, in contrast to the matter–antimatter symmetry that occurs naturally in the Standard Model. According to astrophysical and cosmological evidence discussed in Chapter 13, ordinary baryonic matter constitutes only 5% of the universe. There is an expectation that answers, or at least some initial directions, will be uncovered by the experimental programme of the LHC and future neutrino experiments.

1.7 Chapter outline

Physics is an experimental science, and we would not have our current understanding of particle physics without the use of advanced particle accelerators and detectors. Chapter 3 introduces particle accelerators and explains some of the critical technology required for the successful operation of the LHC. Chapter 4 gives an introduction to the fundamental physics of particle detectors, with an emphasis on the modern techniques used at the LHC. This is obviously a vital subject for particle physics and the chapter attempts to describe this in greater depth than conventional undergraduate textbooks. The applications of these principles to particular experiments will be described in other chapters.

Some of the theoretical and mathematical concepts such as symmetries that will be required throughout the rest of the book are covered in Chapter 2. An introduction to relativistic quantum mechanics is given in Chapter 6. The static quark model for hadrons is described in Chapter 5. The use of scattering experiments to probe the dynamic nature of quarks is covered in Chapter 9, which gives an outline of the quark–parton model as well as the evidence for gluons and a brief introduction to quantum chromodynamics (QCD). The weak interaction is introduced in Chapter 7, starting with the weak interaction of leptons. This is then extended to include quarks and the chapter ends with an introduction to electroweak (EW) unification, including the prediction of the W and Z bosons. A wide range of experiments that support EW unification are covered in Chapter 8, particularly those made possible by the high energies of the LEP, Tevatron, and LHC. Flavour oscillations and CP

violation in the quark sector are explained in Chapter 10. Similar oscillations are seen in the neutrino sector, and the formalism and key experimental results are described in Chapter 11. The intriguing possibility that CP violation in neutrino oscillations could explain the observed matter–antimatter asymmetry in the universe is briefly reviewed. The Higgs mechanism is a fundamental aspect of the Standard Model. The theory and experimental evidence for the existence of a Higgs boson are given in Chapter 12. Finally, Chapter 13 starts with a review of Standard Model physics at the LHC; it then explains why this is not the end of the story. Although the Standard Model is remarkably successful in explaining the current LHC data, there remain compelling reasons to believe that it is an incomplete theory. These are outlined in this chapter, together with a discussion of theoretical ideas beyond the Standard Model that might cure some of its problems. The evidence for dark matter and dark energy is also reviewed, as well as the different attempts to discover dark matter.

1.8 How to read this book

As we mentioned in the Preface, we have the ambitious aim of covering all aspects of particle physics. This inevitably means that not all topics will be of equal interest to all readers. The next three chapters provide technical information: Chapter 2 on mathematical methods; Chapters 3 and 4 on accelerators and detectors, respectively. Depending on the reader’s interest or experience, these can be skipped or returned to later. Similarly, Chapter 6, an introduction to relativistic quantum mechanics and the Dirac equation, is not essential for understanding many of the experimental results, but it is crucial for understanding the concept of antiparticles. Chapter 5 shows how the static quark model can explain the observed pattern of hadronic masses and quantum numbers. The second half of the book (Chapters 7–13) covers all aspects of experimental particle physics, informed by recent results from the LHC and other particle accelerators and experiments. If you need to use the most recent and accurate experimental results, you should consult the PDG tables [115].

Chapter summary

- Aim of the book—a practical introduction to particle physics.
- The basic building blocks—leptons and quarks.
- The electroweak force and the photon, W^\pm , and Z^0 .
- The strong force, QCD, and gluons
- Accelerators, colliders, and detectors

Further reading

- Close, F., Marten, M., and Sutton, C. (1987). *The Particle Explosion*. Oxford University Press.
- Pais, A. (1986). *Inward Bound: Of Matter and Forces in the Physical World*. Oxford University Press.
- Feynman, R. P. and Weinberg, S. (1987). *Elementary Particles and the Laws of Physics: The 1986 Dirac Memorial Lectures*. Cambridge University Press.

Mathematical methods

2

This chapter covers rotational and Lorentz invariance. Related space-time symmetries such as parity, time reversal, and charge conjugation are also defined. We give a brief introduction to group theory and its use in the mathematical construction of the standard model—both in the electroweak sector and in the strong interaction. The idea and usefulness of approximate internal symmetries are explored, using nuclear isospin as an example. This chapter also covers the essential steps in connecting calculations to measured quantities such as cross sections and decay rates. It is assumed that the reader is familiar with the quantization of angular momentum in non-relativistic quantum mechanics and has taken a first course in special relativity.

While investigating the invariance of general relativity in 1918, Emily Noether determined the conserved quantities for all physical laws that are based on a continuous symmetry. Specifically, there are the following associations between symmetries and conserved quantities:

Symmetry		Conserved quantity
spatial rotation	↔	angular momentum
temporal translation	↔	energy
spatial translation	↔	momentum
electromagnetic gauge invariance	↔	electric charge

Three discrete symmetries are also important in nuclear and particle physics: spatial parity (P), charge conjugation (C), and time reversal (T). All three are good symmetries of both the electromagnetic and strong interactions. The weak interaction famously breaks both C and P symmetries maximally but is CP -invariant for many processes. Violation of CP invariance has been observed in the interactions of neutral meson systems, particularly kaons and beauty mesons.¹ The product of all three, CPT , is expected to be a universal symmetry of physics and is a cornerstone of quantum field theory.

2.1 Discrete symmetries	15
2.2 Addition of angular momentum	19
2.3 Spatial rotations	23
2.4 Lorentz invariance	29
2.5 Transitions and observables	31
2.6 Luminosity and event rates	42
2.7 Group theory	43
Chapter summary	46
Further reading	46
Exercises	47

¹ CP violation and its consequences for neutral mesons are covered in Chapter 10.

2.1 Discrete symmetries

2.1.1 Spatial parity

The parity operator performs a spatial inversion through the origin:

$$\psi'(\mathbf{x}, t) = P\psi(\mathbf{x}, t) = \psi(-\mathbf{x}, t)$$

Applying the parity operator twice must return the original state:

$$PP\psi(\mathbf{x}, t) = \psi(\mathbf{x}, t), \quad \text{so} \quad P^2 = 1$$

To preserve the normalization of the wavefunction,

$$\begin{aligned} \langle \psi | \psi \rangle &= \langle \psi' | \psi' \rangle \\ &= \langle \psi | P^\dagger P | \psi \rangle \end{aligned}$$

Therefore,

$$\begin{aligned} P^\dagger P &= I && (P \text{ is unitary}) \\ \text{and since } P^2 &= 1, && P^\dagger = P && (P \text{ is Hermitian}) \end{aligned}$$

which implies that parity can be an observable with eigenvalues ± 1 .

Parity changes the direction of vector quantities (\mathbf{r} , \mathbf{p}) but conserves quantities that are products of vectors, such as $\mathbf{j} = \mathbf{r} \times \mathbf{p}$. Furthermore, since P has no explicit time dependence,

$$i \frac{d\langle P \rangle}{dt} = [P, H].$$

Therefore, parity is a constant of the motion (i.e. **conserved**) when the interaction Hamiltonian commutes with P .

We define the following **intrinsic parity** of fundamental particles:

- **Spin-1 bosons:** Gluons and the photon have intrinsic parity $P = -1$:

$$P|\gamma\rangle = -|\gamma\rangle \quad \text{and} \quad P|g\rangle = -|g\rangle$$

- **Spin- $\frac{1}{2}$ fermions:** particles are of opposite parity to antiparticles—this follows from the Dirac equation (see Chapter 6). The conventional choice is

$$\begin{aligned} P|e^-\rangle &= P|\nu\rangle = P|q\rangle = +1 \\ P|e^+\rangle &= P|\bar{\nu}\rangle = P|\bar{q}\rangle = -1 \end{aligned}$$

The weak gauge bosons (Z^0 and W^\pm) are not eigenstates of spatial parity and thus do not have a definite parity quantum number.

2.1.2 Charge conjugation

The charge-conjugation operator C changes a particle into its antiparticle. Generally, few particles are eigenstates of C , for example a u quark has charge $+\frac{2}{3}$, its antiparticle $-\frac{2}{3}$. Nevertheless, C is a useful quantity when considering electromagnetic or strong decays of neutral mesons.

The photon is a neutral, fundamental particle. Its intrinsic charge quantum number can be inferred by considering its correspondence with classical wave theory. It is clear that upon reversing charge, the electric field and the electromagnetic scalar potential change sign:

$$\mathbf{E}(\mathbf{x}, t) \rightarrow -\mathbf{E}(\mathbf{x}, t), \quad \phi(\mathbf{x}, t) \rightarrow -\phi(\mathbf{x}, t)$$

The vector potential $\mathbf{A}(\mathbf{x}, t)$, which is connected with the photon wavefunction, is related to \mathbf{E} and ϕ by

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}$$

Inserting the charge-reversed \mathbf{E} and ϕ requires

$$C\mathbf{A} = -\mathbf{A}$$

It is then simple to identify $C|\pi^0\rangle = +1|\pi^0\rangle$ from the dominant electromagnetic decay $\pi^0 \rightarrow \gamma\gamma$ (branching ratio = 98.8%).

2.1.3 Time reversal

In line with the spatial parity transformation, we might expect that time reversal would be given by

$$\psi_T(\mathbf{x}, t) = T\psi(\mathbf{x}, t) = \eta_T\psi(\mathbf{x}, -t), \quad \text{where } |\eta_T| = 1$$

Both classical mechanics and electromagnetism respect time reversal. In classical mechanics, for a time-independent potential $V(x)$, Newton's equations of motion can be derived from the energy-conservation equation²

$$E = \frac{1}{2}m\dot{x}^2 + V(x)$$

by differentiation with respect to time, giving

$$m\ddot{x}\dot{x} + \frac{dV}{dx}\dot{x} = 0, \quad \text{or} \quad m\frac{d^2x}{dt^2} = -\frac{dV}{dx}$$

so the equation of motion is unchanged by the change $t \rightarrow -t$. However, this will not be correct for quantum mechanics, because Schrödinger's equation involves a first-order time derivative. For a time-independent Hamiltonian \hat{H} and ψ an eigenstate of energy,

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi = E\psi \quad \implies \quad \psi(x, t) = \psi(x, 0)e^{-iEt/\hbar}$$

If we apply the T operator as defined above to this equation, we find

$$T(\psi(x, t)) = \eta_T\psi(x, 0)e^{+iEt/\hbar}, \quad \text{where } |\eta_T| = 1$$

The time-reversed state appears to have negative energy.³ What matters in quantum mechanics is what it predicts for observable quantities, and this requires calculating normalized matrix elements, for which we need $\psi^*(x, t) = \psi^*(x, 0)e^{+iEt/\hbar}$ as well as $\psi(x, t)$. Looking at the time dependence of this state gives a hint as to how to proceed: we modify the T operator so that, in addition to requiring the change $(\mathbf{x}, t) \rightarrow (\mathbf{x}, -t)$, we demand that $\psi \rightarrow \psi^*$, so now

$$\psi_T(\mathbf{x}, t) \equiv \psi_T(x, t) = \eta_T\psi^*(\mathbf{x}, -t) = \eta_T\psi^*(\mathbf{x}, 0)e^{-iEt/\hbar}$$

²We work in one spatial dimension for simplicity.

³We will consider another view of negative-energy states in Chapter 6.

Such an operator is known as *anti-unitary*. It is straightforward to show that the normalization of $\psi(\mathbf{x}, t)$ is unchanged by the T operation, provided that $|\eta_T|^2 = 1$.

2.1.4 J^{PC} of hadrons

Parity and, where applicable, charge quantum numbers are often quoted, for a particular state, with its total angular momentum $J = L + S$ as a J^{PC} number. The J^P of a particle is closely related to the spatial transformation properties of the state. Particles with $J^P = 0^-$ are called *pseudoscalar* particles and those with $J^P = 0^+$ *scalar*. Particles with $J^P = 1^-$ are called *vector* particles and those with $J^P = 1^+$ *axial vector*.

Mesons

Mesons are $q\bar{q}$ bound states. As these have opposite parity, a ground-state meson will always have $P = -1$. Excited states bring additional parity factors according to $(-1)^L$:

$$P_{\text{meson}} = (-1)^{L+1}$$

C is defined (for light neutral mesons only) by interchanging $q \leftrightarrow \bar{q}$ and swapping their positions and spin:

$$C_{\text{meson}} = (-1)^{L+S}$$

Baryons

Baryons contain three quarks and as such cannot be their own antiparticle; C is undefined. The calculation of baryon parity is more complex than for mesons because one must consider the angular momentum of a three-body system. The intrinsic parity of a baryon is $(+1)^3 = +1$; similarly, it is -1 for an antibaryon. In full,

$$P_{\text{baryon}} = B(-1)^{L_{12}}(-1)^{L_3}$$

where B is the baryon number, L_{12} is the angular momentum between quarks 1 and 2, and L_3 is the angular momentum of the third quark relative to the 1-2 pair.

2.1.5 Useful examples

- *What is the J^{PC} of a π^0 ?*

The π^0 is in the ground state, with $L = S = J = 0$. From the formula, $P = -1$ and $C = +1$. So $J^{PC} = 0^{-+}$.

- *Is $\rho^0 \rightarrow \pi^+\pi^-$ an allowed decay?*

This is an excited initial state, with $S = J = 1$, $L = 0$, so $P = -1$. Also, from above, $C = -1$, so $J^{PC} = 1^{--}$.

The pair of pions have intrinsic parity $+1$ from $(-1)_{\pi^+}(-1)_{\pi^-}$ and each $J = 0$.

So producing them in an $L = 1$ state will simultaneously conserve J , P , and C .

- What about $\rho^0 \rightarrow \pi^0\pi^0$?

This would be similar to the previous example, requiring the pion pair to be in an $L = 1$ state.

However, applying C to the final state has no effect:⁴

$$C|\pi^0\pi^0\rangle = |\pi^0\pi^0\rangle, \quad C = +1, \quad J^{PC}(\pi^0\pi^0) = 1^{-+}$$

So this process is forbidden by charge violation with the strong or electromagnetic force.

Indeed, it has never been observed.

- $\rho^0 \rightarrow \ell^+\ell^-$

This is similar to $\rho^0 \rightarrow \pi^+\pi^-$ except that the $\ell^+\ell^-$ system has intrinsic parity -1 from $(+1)_{\ell^-}(-1)_{\ell^+}$. Producing a final state with the spins aligned, $S = 1$, simultaneously conserves J , P , and C . Note that this process only proceeds via the electromagnetic force and is therefore only $\sim 10^{-4}$ as likely as the $\pi^+\pi^-$ mode.

- $\rho^0 \rightarrow \pi^0\gamma$

The initial state is $J^{PC} = 1^{--}$. The photon has $J^{PC} = 1^{--}$ and the π^0 has 0^{-+} .

But, in addition to their intrinsic parity, photons carry parity $(-1)^L$ away from a system.

Therefore, this electromagnetic decay is allowed.

- What is the J^{PC} of a K^+

$|\bar{s}u\rangle$ is not an eigenstate of charge inversion, so the C quantum number is undefined.

Therefore, the label becomes $J^P = 0^-$ (ground-state kaon).

⁴Applying charge conjugation to a single neutral pion gives $C|\pi^0\rangle = |\pi^0\rangle$, therefore applying the charge conjugation operator to two neutral pions gives a factor of one. Note that this decay mode is also forbidden by Bose-Einstein symmetry: the two pions must be in an $L = 1$ state to conserve angular momentum but this would require the wave function to be anti-symmetric with respect to exchange of identical bosons.

2.2 Addition of angular momentum

2.2.1 Angular momentum in quantum mechanics

In quantum mechanics, the angular momentum operators J_x , J_y , J_z do not commute, but they do satisfy the commutation relations

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (2.1)$$

where $\epsilon_{ijk} = +1$ if ijk is an even permutation of xyz (e.g. yzx), -1 if it is an odd permutation (e.g. zyx), and 0 if any of the indices are identical (e.g. xyx). The only operator that commutes with J_x , J_y , J_z is the total angular momentum J^2 :

$$J^2 = J_x^2 + J_y^2 + J_z^2, \quad \text{and} \quad [J^2, J_i] = 0 \quad (i = x, y, z)$$

Eigenfunctions of J^2 and J_z are labelled with the eigenvalues of J^2 and J_z :

$$\begin{aligned} J^2|j, m\rangle &= j(j+1)|j, m\rangle \\ J_z|j, m\rangle &= m|j, m\rangle \end{aligned}$$

where $m \in \{-j, j+1, \dots, j-1, j\}$. So, for a given j value there are $2j+1$ values of J_z , with the states related by raising and lowering operators

$$J_{\pm} = J_x \pm iJ_y \quad (2.2)$$

which, from eqn 2.1, satisfy

$$[J_z, J_{\pm}] = \pm J_{\pm} \quad (2.3)$$

Using this commutation relation, we have

$$\begin{aligned} J_z J_- |j, m\rangle &= (J_- J_z - J_-) |j, m\rangle \\ &= J_- (J_z - 1) |j, m\rangle \\ &= (m - 1) J_- |j, m\rangle \end{aligned}$$

Similarly, $m + 1$ is the eigenvalue of J_z applied to the ‘raised’ state $J_+ |j, m\rangle$. So we can generally write

$$J_{\pm} |j, m\rangle = \mathbf{C}_{\pm}(j, m) |j, m \pm 1\rangle \quad (2.4)$$

with the boundary condition that \mathbf{C}_{\pm} is required to be 0 for $J_- |j, -j\rangle$ and $J_+ |j, +j\rangle$.

To derive these constants, we note that J_+ and J_- are Hermitian adjoints (although they are not Hermitian operators):

$$\begin{aligned} \langle j, m | J_+^{\dagger} |j, m + 1\rangle &= \langle j, m | J_- |j, m + 1\rangle \\ &= \mathbf{C}_-(j, m + 1) \langle j, m | j, m\rangle \\ &= \mathbf{C}_-(j, m + 1) \end{aligned} \quad (2.5)$$

We take the complex conjugate of this equation to give

$$\langle j, m + 1 | J_+ |j, m\rangle = \mathbf{C}_+^*(j, m) \langle j, m + 1 | j, m + 1\rangle = \mathbf{C}_+^*(j, m)$$

so the *lowering* coefficient of the $m + 1$ state is the same as the *raising* coefficient of the m state:

$$\mathbf{C}_-(j, m + 1) = \mathbf{C}_+^*(j, m) = \mathbf{C} \quad (2.6)$$

Therefore, applying both operators successively,

$$J_- J_+ |j, m\rangle = \mathbf{C}^2 |j, m\rangle \quad (2.7)$$

where the double operator can be broken down to

$$\begin{aligned} J_- J_+ &= J_x^2 + J_y^2 + i(J_x J_y - J_y J_x) \\ &= J_x^2 + J_y^2 + J_z^2 - J_z^2 + i[J_x, J_y] \\ &= J^2 - J_z^2 - J_z \\ &= J^2 - J_z(J_z + 1) \end{aligned} \quad (2.8)$$

from which we can easily identify the eigenvalues of C^2 and the coefficients C_{\pm}

$$\begin{aligned} C^2 &= j(j+1) - m(m+1) \\ C_{-}(j, m) &= \sqrt{j(j+1) - m(m-1)} \\ C_{+}(j, m) &= \sqrt{j(j+1) - m(m+1)} \end{aligned} \quad (2.9)$$

2.2.2 Addition and Clebsch–Gordan coefficients

Let J_1 and J_2 be two angular momentum operators, for example the orbital angular momentum and spin of a particle:

$$\begin{aligned} J_1^2 |j_1, m_1\rangle &= j_1(j_1+1)|j_1, m_1\rangle \\ J_2^2 |j_2, m_2\rangle &= j_2(j_2+1)|j_2, m_2\rangle \\ J_{1z} |j_1, m_1\rangle &= m_1|j_1, m_1\rangle \\ J_{2z} |j_2, m_2\rangle &= m_2|j_2, m_2\rangle \end{aligned} \quad (2.10)$$

We define the combined operators

$$\begin{aligned} J^2 &= (J_{1x} + J_{2x})^2 + (J_{1y} + J_{2y})^2 + (J_{1z} + J_{2z})^2 \\ J_z &= J_{1z} + J_{2z} \end{aligned}$$

acting on

$$\psi = |j_1, m_1\rangle |j_2, m_2\rangle$$

ψ is an eigenfunction of J_z with eigenvalue $m = m_1 + m_2$ but it is generally not an eigenfunction of J^2 . However, linear combinations of ψ can produce eigenfunctions $\Psi(j, m, j_1, j_2)$ of J^2 with eigenvalues $j(j+1)$:

$$\Psi = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} C_{j_1 j_2}(j, m, m_1, m_2) \psi \quad (2.11)$$

The coefficients $C_{j_1 j_2}(j, m, m_1, m_2)$ are the Clebsch–Gordan⁵ coefficients. They are the probability amplitudes that we measure, from a combined system of $|j_1, m_1\rangle |j_2, m_2\rangle$, with a combined angular momentum of $j(j+1)$ when applying the operator J^2 . The details can be found in quantum mechanics texts; here we will consider a useful example.

⁵Clebsch and Gordan were nineteenth-century mathematicians who identified these coefficients during the development of the theory of Lie algebras.

2.2.3 Calculation of Clebsch–Gordan coefficients

Consider two particles $|j_1, m_1\rangle$ and $|j_2, m_2\rangle$ forming a combined state $|j, m\rangle$, where $j_1 = 1$ and $j_2 = \frac{1}{2}$. Evidently, j can be $\frac{1}{2}$ or $\frac{3}{2}$. The following two points are key:

- Coefficients should normalize to 1.
- On raising (lowering) a state to an m higher (lower) than J ($-J$), the resultant state = 0.

$\frac{3}{2}$ states:

Evidently, the states with maximum ($\frac{3}{2}$) and minimum ($-\frac{3}{2}$) m are

$$\begin{aligned} \left| \frac{3}{2}, \frac{3}{2} \right\rangle &= |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle \\ \left| \frac{3}{2}, -\frac{3}{2} \right\rangle &= |1, -1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \end{aligned}$$

Using the definition $\mathbf{C}_{\pm}(j, m) = \sqrt{j(j+1) - m(m \pm 1)}$, we have

$$\begin{aligned} J_- \left| \frac{1}{2}, \frac{1}{2} \right\rangle &= 1 \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \\ J_- |1, 1\rangle &= \sqrt{2} |1, 0\rangle \\ J_- |1, 0\rangle &= \sqrt{2} |1, -1\rangle \\ J_- |1, -1\rangle &= 0 \end{aligned}$$

Now operating on the combined state, we have

$$\begin{aligned} J_- \left| \frac{3}{2}, \frac{3}{2} \right\rangle &= J_- |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle \\ \sqrt{3} \left| \frac{3}{2}, \frac{1}{2} \right\rangle &= \sqrt{2} |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle + |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \\ \left| \frac{3}{2}, \frac{1}{2} \right\rangle &= \sqrt{\frac{2}{3}} |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle + \sqrt{\frac{1}{3}} |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \end{aligned}$$

In an analogous way,

$$\left| \frac{3}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{2}{3}} |1, 0\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \sqrt{\frac{1}{3}} |1, -1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle$$

$\frac{1}{2}$ states:

$$\left| \frac{1}{2}, \frac{1}{2} \right\rangle = a |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + b |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle$$

with $a^2 + b^2 = 1$ because of normalization. We now apply J_+ :

$$\begin{aligned} J_+ \left| \frac{1}{2}, \frac{1}{2} \right\rangle &= 0 \\ &= a |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle + \sqrt{2} b |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle \end{aligned}$$

Therefore, $a + \sqrt{2}b = 0$, which with $a^2 + b^2 = 1$ gives $a = \sqrt{\frac{2}{3}}$ and $b = -\sqrt{\frac{1}{3}}$.

These results are summarized in Tables 2.1 and 2.2.

$1 \otimes \frac{1}{2}$	j	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{3}{2}$
	m	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$
m_1	m_2						
+1	$+\frac{1}{2}$	1	0	0	0	0	0
+1	$-\frac{1}{2}$	0	$\sqrt{\frac{1}{3}}$	$\sqrt{\frac{2}{3}}$	0	0	0
0	$+\frac{1}{2}$	0	$\sqrt{\frac{2}{3}}$	$-\sqrt{\frac{1}{3}}$	0	0	0
0	$-\frac{1}{2}$	0	0	0	$\sqrt{\frac{1}{3}}$	$\sqrt{\frac{2}{3}}$	0
-1	$+\frac{1}{2}$	0	0	0	$-\sqrt{\frac{2}{3}}$	$\sqrt{\frac{1}{3}}$	0
-1	$-\frac{1}{2}$	0	0	0	0	0	1

Table 2.1 Clebsch–Gordan coefficients for $j_1 = 1$ and $j_2 = \frac{1}{2}$.

$\frac{1}{2} \otimes \frac{1}{2}$	j	1	1	0	1
	m	1	0	0	-1
m_1	m_2				
$+\frac{1}{2}$	$+\frac{1}{2}$	1	0	0	0
$+\frac{1}{2}$	$-\frac{1}{2}$	0	$\sqrt{\frac{1}{2}}$	$\sqrt{\frac{1}{2}}$	0
$-\frac{1}{2}$	$+\frac{1}{2}$	0	$\sqrt{\frac{1}{2}}$	$-\sqrt{\frac{1}{2}}$	0
$-\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	1

Table 2.2 Clebsch–Gordan coefficients for $j_1 = \frac{1}{2}$ and $j_2 = \frac{1}{2}$.

2.3 Spatial rotations

Consider a small rotation ϵ about the y axis,

$$\mathbf{x}' = \mathbf{R}_y(\epsilon)\mathbf{x}$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \epsilon & 0 & \sin \epsilon \\ 0 & 1 & 0 \\ -\sin \epsilon & 0 & \cos \epsilon \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

and its inverse,

$$\mathbf{x} = \mathbf{R}_y^{-1}(\epsilon)\mathbf{x}'$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos \epsilon & 0 & -\sin \epsilon \\ 0 & 1 & 0 \\ \sin \epsilon & 0 & \cos \epsilon \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

We impose invariance:

$$\psi'(\mathbf{x}') = \psi(\mathbf{x}) = \psi(\mathbf{R}_y^{-1}(\epsilon)\mathbf{x}')$$

Without loss of generality, we take a specific point $\mathbf{x}' = \mathbf{a}$ and find a relation between $\psi'(\mathbf{a})$ and $\psi(\mathbf{a})$:

$$\begin{aligned} \psi'(\mathbf{a}) &= \psi(\mathbf{R}_y^{-1}(\epsilon)\mathbf{a}) \\ &= \psi(a_x \cos \epsilon - a_z \sin \epsilon, a_y, a_z \cos \epsilon + a_x \sin \epsilon) \\ &\rightarrow \psi(a_x - \epsilon a_z, a_y, a_z + \epsilon a_x) \quad \text{as } \epsilon \rightarrow 0 \\ &= \psi(\mathbf{a}) + \epsilon a_x \frac{\partial \psi(\mathbf{a})}{\partial z} - \epsilon a_z \frac{\partial \psi(\mathbf{a})}{\partial x} + \dots \quad (\text{Taylor expansion}) \\ &\approx \psi(\mathbf{a}) \left[1 + \epsilon \left(a_x \frac{\partial}{\partial z} - a_z \frac{\partial}{\partial x} \right) \right] \\ &= \underbrace{(1 - i\epsilon J_y)}_{U_r(\epsilon)} \psi(\mathbf{a}) \quad \left[\text{since } J_y = -i \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \right] \end{aligned}$$

Conservation \leftrightarrow invariance

⁶Here we explicitly assume that the non-relativistic Schrödinger equation is valid. The resulting conservation law is therefore only valid in the non-relativistic limit. The relativistic case requires the use of the Dirac equation and this will be discussed in chapter 6

Consider the time variation of U_r :⁶

$$\begin{aligned} &\frac{d}{dt} \langle \phi(t) | U_r | \varphi(t) \rangle \\ &= \left[\frac{d}{dt} \langle \phi(t) | \right] U_r | \varphi(t) \rangle + \langle \phi(t) | \frac{dU_r}{dt} | \varphi(t) \rangle + \langle \phi(t) | U_r \left[\frac{d}{dt} | \varphi(t) \rangle \right] \\ &= \langle \phi(t) | \frac{dU_r}{dt} | \varphi(t) \rangle + i \langle \phi(t) | U_r H - H U_r | \varphi(t) \rangle \end{aligned}$$

Hence, U_r is invariant if J_y (the tricky part of U_r) is independent of time and commutes with the Hamiltonian; i.e. the eigenvalues of $U_r(\epsilon)$ are constant. Angular momentum is conserved owing to the requirement that the wavefunction be invariant under rotation.

⁷Section 2.7 covers the essentials of group theory.

Rotation is a Lie group⁷ and so any rotation can be expressed as the successive application of the infinitesimal rotation:

$$\begin{aligned} U_r(\beta) &= \lim_{n \rightarrow \infty} \left(1 - i \frac{\beta}{n} J_y \right)^n \\ &= e^{-i\beta J_y} \end{aligned} \tag{2.12}$$

or, for a general three-dimensional rotation,

$$U_{\mathbf{r}}(\boldsymbol{\theta}) = e^{-i\boldsymbol{\theta}\cdot\mathbf{J}}$$

In the language of operators, the angular momentum operator J_y is said to be the *generator* of rotations about the y axis.

Euler angles

Generic rotations can be parameterized by Euler angles, which are defined by three successive rotations:

- (1) rotation by an angle γ about the z axis;
- (2) rotation by an angle β about the original y axis;
- (3) rotation by an angle α about the new z' axis.

This is inconvenient, because these definitions use two different bases. However, this transformation is actually the same as

- (1) rotation by an angle α about the *original* z axis;
- (2) rotation by an angle β about the *original* y axis;
- (3) rotation by an angle γ about the *original* z axis.

The generic rotation of wavefunctions can be represented by D -matrices:

$$\begin{aligned} D_{m',m}^j(\alpha\beta\gamma) &= \langle j, m' | e^{-i\alpha J_z} e^{-i\beta J_y} e^{-i\gamma J_z} | j, m \rangle \\ &= e^{-i(\alpha m' + \gamma m)} d_{m',m}^j(\beta) \end{aligned} \quad (2.13)$$

Rotation matrix: $d_{m',m}^j(\beta)$

Although the y projection is unchanged, the z direction has changed, so the quantum number m is not the same—it is now projected onto a new z' axis. A state $|j, m\rangle$ transforms under a rotation β about the y axis into a linear combination on the $2j + 1$ states $|j, m'\rangle$:

$$e^{-i\beta J_y} |j, m\rangle = \sum_{m'} d_{m',m}^j(\beta) |j, m'\rangle \quad (2.14)$$

Multiplying with $\langle j, m' |$ gives

$$d_{m',m}^j(\beta) = \langle j, m' | e^{-i\beta J_y} |j, m\rangle \quad (2.15)$$

Calculation of the matrix proceeds as follows:

- (1) From inspection of the expansion of $e^{-i\beta J_y}$,

$$e^{-i\beta J_y} = 1 - i\beta J_y - \frac{1}{2!}\beta^2 J_y^2 + \frac{i}{3!}\beta^3 J_y^3 + \frac{1}{4!}\beta^4 J_y^4 + \dots$$

- (2) Look separately for solutions of $J_y^{2n+1}|j, m\rangle$ and $J_y^{2n}|j, m\rangle$.

- (3) Recall the raising/lowering operators $J_{\pm} = J_x \pm iJ_y$ and the Clebsch–Gordan coefficients $C_{\pm}^{jm} = \sqrt{j(j+1) - m(m \pm 1)}$. Then

$$\begin{aligned} J_+ - J_- &= J_x + iJ_y - J_x + iJ_y \\ J_y &= -\frac{i}{2}(J_+ - J_-) \end{aligned}$$

so

$$\begin{aligned} J_y|1, 1\rangle &= -\frac{i}{2} \left(\cancel{C_+^{11}} \cdot 0 - C_-^{11}|1, 0\rangle \right) \quad \text{with } C_-^{11} = \sqrt{2} \\ &= \frac{i}{\sqrt{2}}|1, 0\rangle \end{aligned}$$

- (4) Operate again with J_y :

$$\begin{aligned} J_y^2|1, 1\rangle &= \frac{i}{\sqrt{2}}J_y|1, 0\rangle \\ &= \frac{1}{2\sqrt{2}}(C_+^{10}|1, 1\rangle - C_-^{10}|1, -1\rangle) \\ &= \frac{1}{2}(|1, 1\rangle - |1, -1\rangle), \quad \text{since } C_+^{10} = C_-^{10} = \sqrt{2} \end{aligned}$$

- (5) And again:

$$\begin{aligned} J_y^3|1, 1\rangle &= \frac{1}{2}J_y(|1, 1\rangle - |1, -1\rangle) \\ &= \frac{1}{2} \left[\frac{-i}{2}(J_+ - J_-)|1, 1\rangle - \frac{-i}{2}(J_+ - J_-)|1, -1\rangle \right] \\ &= \frac{i}{4} \left(\sqrt{2}|1, 0\rangle + \sqrt{2}|1, 0\rangle \right) \\ &= \frac{i}{\sqrt{2}}|1, 0\rangle \\ &= J_y|1, 1\rangle \end{aligned}$$

- (6) Note the cyclical pattern and conclude that

$$\begin{aligned} J_y^{2n+1}|1, 1\rangle &= \frac{i}{\sqrt{2}}|1, 0\rangle \\ J_y^{2n}|1, 1\rangle &= \frac{1}{2}(|1, 1\rangle - |1, -1\rangle) \end{aligned}$$

- (7) Which then leads, for each specific $\langle j, m' |$ state, to the following:

$$\begin{aligned} \langle 1, 1 | J_y^{2n+1} | 1, 1 \rangle &= 0, & \langle 1, 1 | J_y^{2n} | 1, 1 \rangle &= \frac{1}{2} \\ \langle 1, -1 | J_y^{2n+1} | 1, 1 \rangle &= 0, & \langle 1, -1 | J_y^{2n} | 1, 1 \rangle &= -\frac{1}{2} \\ \langle 1, 0 | J_y^{2n+1} | 1, 1 \rangle &= \frac{i}{\sqrt{2}}, & \langle 1, 0 | J_y^{2n} | 1, 1 \rangle &= 0 \end{aligned}$$

(8) With these relations, the $d_{m'm}^j$ coefficients are calculated as

$$\begin{aligned}
 d_{11}^1 &= \langle 1, 1 | e^{-i\beta J_y} | 1, 1 \rangle \\
 &= \langle 1, 1 | 1 + \cancel{(-i\beta J_y)} + \frac{1}{2!}(-i\beta J_y)^2 + \frac{1}{3!}\cancel{(-i\beta J_y)^3} \\
 &\quad + \frac{1}{4!}(-i\beta J_y)^4 + \dots | 1, 1 \rangle \\
 &= 1 - \frac{1}{2!} \frac{\beta^2}{2} + \frac{i}{4!} \frac{\beta^4}{2} - \dots \\
 &= \frac{1}{2} \left[1 + \left(1 - \frac{1}{2!} \beta^2 + \frac{1}{4!} \beta^4 - \dots \right) \right] \\
 &= \frac{1}{2} (1 + \cos \beta)
 \end{aligned}$$

$$\begin{aligned}
 d_{-11}^1 &= \langle 1, -1 | e^{-i\beta J_y} | 1, 1 \rangle \\
 &= \frac{1}{2} \left[1 - \left(1 - \frac{1}{2!} \beta^2 + \frac{1}{4!} \beta^4 - \dots \right) \right] \quad (\langle 1, -1 | 1, 1 \rangle = 0 \text{ of course}) \\
 &= \frac{1}{2} (1 - \cos \beta)
 \end{aligned}$$

$$\begin{aligned}
 d_{01}^1 &= \langle 1, 0 | e^{-i\beta J_y} | 1, 1 \rangle \\
 &= \langle 1, 0 | -\frac{i}{\sqrt{2}} i\beta + \frac{i}{\sqrt{2}} \frac{1}{3!} \beta^3 + \dots | 1, 0 \rangle \\
 &= \frac{1}{\sqrt{2}} \left(\beta - \frac{1}{3!} \beta^3 + \dots \right) \\
 &= \frac{1}{\sqrt{2}} \sin \beta
 \end{aligned}$$

$$\begin{aligned}
 d_{00}^1 &= \langle 1, 0 | e^{-i\beta J_y} | 1, 0 \rangle \\
 &= -i\sqrt{2} \langle 1, 0 | e^{-i\beta J_y} J_y | 1, 1 \rangle \\
 &= \cos \beta
 \end{aligned}$$

Example: $e^- e^+ \rightarrow \mu^- \mu^+$

With reference to Fig. 2.1, the incoming left-handed electron annihilates with the right-handed positron. In the electromagnetic interaction, a photon is exchanged with the outgoing muon pair. As will be discussed in detail in Chapter 6, at high energies, helicity⁸ is conserved

⁸Helicity, $\boldsymbol{\sigma} \cdot \mathbf{p}/|\mathbf{p}|$, is the projection of the spin along the momentum direction.

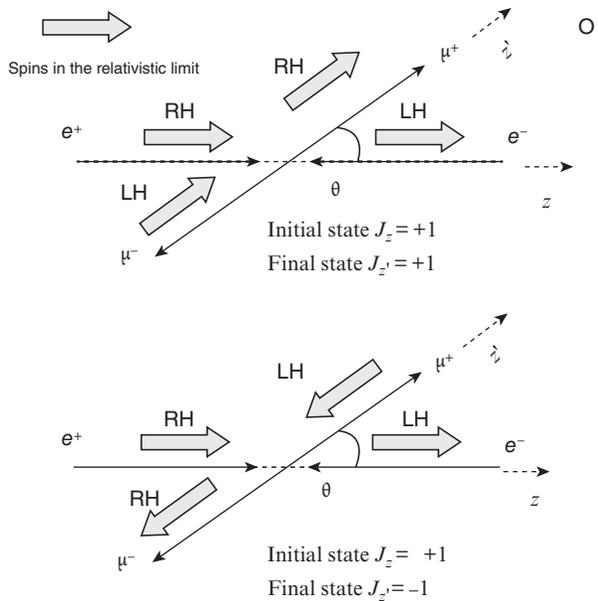


Fig. 2.1 Helicity conservation in $e^+e^- \rightarrow \mu^+\mu^-$.

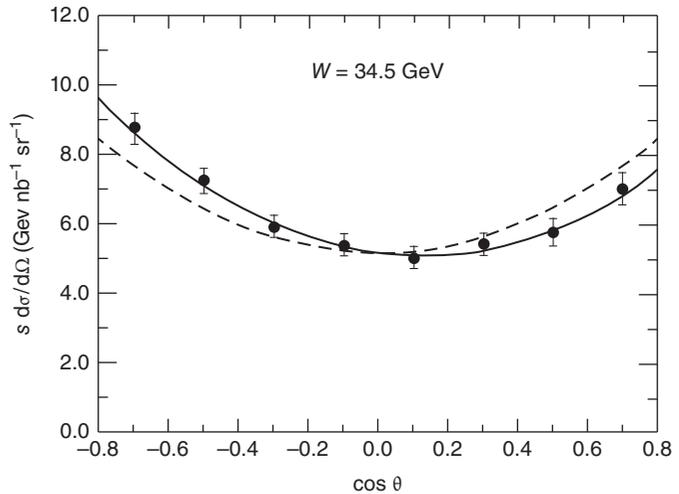


Fig. 2.2 Angular distribution for $e^+e^- \rightarrow \mu^+\mu^-$, TASSO [55].

in this reaction, so the final-state particles must have opposite helicity. The amplitudes are given by

$$\mathcal{A}_{11} = \mathcal{A}(e_L^- e_R^+ \rightarrow \mu_L^- \mu_R^+) \propto d_{1,1}^1 = \frac{1}{2}(1 + \cos \theta)$$

$$\mathcal{A}_{1-1} = \mathcal{A}(e_L^- e_R^+ \rightarrow \mu_R^- \mu_L^+) \propto d_{1,-1}^1 = \frac{1}{2}(1 - \cos \theta)$$

$$\begin{aligned} \frac{d\sigma}{d \cos \theta} &= |\mathcal{A}_{11}|^2 + |\mathcal{A}_{1-1}|^2 \\ &\propto 1 + \cos^2 \theta \end{aligned}$$

The two amplitudes should be of equal intensity because of parity conservation in the electromagnetic interaction. Figure 2.2 shows $e^+e^- \rightarrow \mu^+\mu^-$ data from the TASSO experiment at DESY [55]. It is clear that the angular distribution is not symmetric about $\cos\theta = 0$. This asymmetry is evidence for the off-shell influence of the parity-violating Z^0 interfering with the dominant γ exchange.

2.4 Lorentz invariance

Most high-energy physics requires energy scales $E \gg m_{\text{proton}}$, so it is essential that the requirements of special relativity be respected. In practice, this means identifying suitable 4-vectors and Lorentz invariants. Although the position and direction of particles is important when actually performing experiments, the results are most often derived from knowledge of the energy and momentum of the interacting particles. Here we summarize the essentials and define the Lorentz metric:

- The components of the energy–momentum 4-vector $(E, \mathbf{p}c)$ of a particle of rest mass m are related by

$$E^2 - |\mathbf{p}|^2 = m^2$$

- This relation also defines the Lorentz metric tensor $g_{\mu\nu}$ or $g^{\mu\nu}$:

$$g_{00} = 1, \quad g_{11} = g_{22} = g_{33} = -1, \quad \text{with all other components} = 0$$

- The scalar product of two 4-vectors $X^\mu \equiv (X^0, \mathbf{X})$ and $Y^\mu \equiv (Y^0, \mathbf{Y})$ is

$$X \cdot Y = X_\mu Y^\mu = X^\mu Y_\mu = g_{\mu\nu} X^\mu Y^\nu = g^{\mu\nu} X_\mu Y_\nu$$

The scalar product of two 4-vectors and hence the length of a 4-vector are Lorentz invariants.

Consider next the relationship between the energy–momentum 4-vectors $p = (E, \mathbf{p}c)$ and $p' = (E', \mathbf{p}'c)$ of a particle of mass m in two Lorentz frames S and S' , where S' is moving with a speed $\beta \equiv v/c$ along the z axis in frame S . The p_x and p_y components of 3-momentum, perpendicular to the boost, are unchanged, but the p_z component along the boost direction and the energy are modified:

$$E' = \gamma(E - \beta p_z c) \tag{2.16}$$

$$p'_z c = \gamma(p_z c - \beta E) \tag{2.17}$$

$$p'_x = p_x \tag{2.18}$$

$$p'_y = p_y \tag{2.19}$$

where $\gamma = 1/\sqrt{1-\beta^2}$ and $\beta = v/c$ is the boost. It is convenient to have the transformation expressed in terms of angles with respect to the z axes:

$$E' = \gamma(E - \beta pc \cos \theta) \quad (2.20)$$

$$p' c \cos \theta' = \gamma(pc \cos \theta - \beta E) \quad (2.21)$$

$$p' \sin \theta' = p \sin \theta \quad (2.22)$$

2.4.1 Invariant variables

Many high-energy scattering processes can be measured and analysed in terms of two variables: the centre-of-mass energy E_{CM} and the scattering or production angle of a ‘leading’ final-state particle θ_{CM} . Consider a generic process $a + b \rightarrow c + X$, where a and b are either beam and target or colliding beam particles, c is the leading final-state particle, and X is the, often unmeasured, remainder of the final state. Two invariant variables are useful:

$$s = (p_a + p_b)^2, \quad t = (p_a - p_c)^2 \quad (2.23)$$

where p_a , p_b , and p_c are the 4-momenta of particles a , b , and c . s is the square of the centre-of-mass energy and t is the square of the 4-momentum transfer. For example, consider the process $\pi\pi \rightarrow \pi X$ using a pion beam of energy E_π on a fixed liquid-hydrogen target, for which the energy E'_π and angle θ' of the leading final-state pion are measured. One finds (see Exercise 2.6) that

$$s \equiv E_{\text{CM}}^2 = m_\pi^2 + m_p^2 + 2m_p E_\pi, \quad (2.24)$$

$$t = 2m_\pi^2 - 2(E_\pi E'_\pi - k k' \cos \theta'), \quad (2.25)$$

where k and k' are the magnitudes of the 3-momenta of the two pions.

For a high-energy collider with equal-mass particles (e.g. LEP with e^+e^- or LHC with pp), $s = 4E_{\text{beam}}^2$, ignoring masses. The equivalent fixed-target beam energy required to give the same \sqrt{s} is $E_p \approx s/(2m_p)$; for example to achieve $E_{\text{CM}} = 7$ TeV would require a proton-beam energy of $7^2 \times 10^3 / (2 \times 0.94) \approx 2.6 \times 10^4$ TeV. Clearly, colliders are the most energy-efficient way to reach the highest energies. As already mentioned in Chapter 1 and discussed in more detail in Chapter 3, the key challenge for a collider is to achieve high enough luminosity.

2.4.2 Rapidity

High-energy hadron–hadron interactions tend to produce final states with limited transverse momentum with respect to the initial beam direction. In such circumstances, rapidity y and transverse mass m_{T} are convenient variables, where

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad m_{\text{T}} = \sqrt{m^2 + p_{\text{T}}^2}$$

The 4-momentum p of a particle of mass m , transverse momentum p_T , and an azimuthal angle ϕ ,

$$p = (E, p_T \cos \phi, p_T \sin \phi, p_z)$$

may be described instead as

$$p = (m_T \cosh y, p_T \cos \phi, p_T \sin \phi, m_T \sinh y) \quad (2.26)$$

Rapidity has the approximate range $(\ln(m/2E), \ln(2E/m))$. A difference in rapidity is an invariant under a Lorentz transformation along the beam direction and rapidities are additive under Lorentz boosts along the beam direction. At high energies, when masses can be ignored, y may be approximated by the pseudorapidity η :

$$y \rightarrow \eta \equiv \frac{1}{2} \ln \left(\frac{1 + \cos \theta}{1 - \cos \theta} \right) = -\ln \tan(\theta/2)$$

where θ is the polar angle. The range of pseudorapidity is $(-\infty, \infty)$. Note that $\eta = 0$ (or $y = 0$) is perpendicular to the beam line and large $|\eta|$ (or $|y|$) is close to the beam line. In high-energy hadron–hadron scattering, it is observed that particle production is roughly uniform in units of pseudorapidity.

A useful relation follows from the Jacobian of the transformation from Cartesian to rapidity momentum components:

$$\frac{d^3\mathbf{p}}{E} = p_T dp_T d\phi dy \equiv d^2p_T dy \rightarrow \pi dp_T^2 dy \approx \pi dp_T^2 d\eta \quad (2.27)$$

where for the last two expressions the azimuthal angle has been integrated out.

2.5 Transitions and observables

Particle physics experiments often involve measuring decay rates or scattering cross sections—processes that involve transitions from one state to another. For a particle of total width Γ , the lifetime is given by $\tau = 1/\Gamma$ and the number of particles decays exponentially:

$$N(t) = N(0)e^{-t/\tau} = N(0)e^{-t\Gamma} \quad (2.28)$$

where $N(0)$ and $N(t)$ are the numbers of particles at times 0 and t , respectively.⁹ Often a particle will decay to many final states, so it is useful to define Γ_i , the partial decay rate to final state i . The total rate is then given by $\Gamma = \sum_i \Gamma_i$, where the sum runs over all final states.

Similarly, the attenuation of a particle beam of flux $I(x)$ is given by

$$I(x) = I(0)e^{-x/\ell} \quad (2.29)$$

⁹With $\hbar = c = 1$, both time and length have dimensions of energy⁻¹ and $\tau\Gamma \sim 1$ is the time–energy uncertainty relation.

¹⁰Familiar from the classical kinetic theory of gases as the mean free path.

where ℓ is the collision length¹⁰ given by $\ell = 1/N\sigma$ for a target of number density N and scattering cross section σ . The cross section has physical dimensions of area, so $[\text{energy}]^{-2}$ here. For a thin target of thickness δx , the beam attenuation is given by $\delta I \sim I(0)N\sigma\delta x$.

In general, the S matrix describing a scattering or decay $A \rightarrow B$ is written as $S = 1 + iT$ and the reduced matrix element $\mathcal{M}(B : A)$ is defined by

$$S_{fi} \equiv \langle B | iT | A \rangle = i(2\pi)^4 \delta^4(p_A - p_B) \mathcal{M}(B : A) \quad (2.30)$$

where p_A and p_B are the total 4-momenta of states A and B .

2.5.1 Phase space and decay rates

Transitions to a final state $|B\rangle$ from an initial state $|A\rangle$ are calculated from Fermi's Golden Rule:

$$\Gamma_{fi} = W_{fi} = 2\pi \underbrace{|\mathcal{T}_{fi}|^2}_{\text{dynamics}} \times \underbrace{\rho(E_f)}_{\text{kinematics}} \quad (2.31)$$

where

- Γ_{fi} is the decay width to the final state in question;
- W_{fi} is its transition rate: the number of transitions per unit time;
- \mathcal{T}_{fi} is the matrix element describing the dynamics of the transition:

$$\mathcal{T}_{fi} = \langle B | V | A \rangle \quad (2.32)$$

with $|A\rangle$ and $|B\rangle$ the initial and final states interacting via a potential V ;

- $\rho(E_f)$ is the phase-space factor.

The phase-space factor $\rho(E_f)$

This is the number of states available per unit of energy in the final state. It is important because it connects the physics contained in the matrix element to observable quantities. First, we shall review its calculation using non-relativistic quantum mechanics (NRQM) and explain why this is not appropriate for use in high-energy physics.

Non-relativistic quantum mechanics

In NRQM, the calculation proceeds as follows. Imagine a cube of sides L containing one of the final-state particles with quantized momentum, $p = k$ and $k = 2\pi n$, where n is an integer.¹¹ Then

$$p_x = \frac{2\pi n_x}{L}, \quad p_y = \frac{2\pi n_y}{L}, \quad p_z = \frac{2\pi n_z}{L}$$

$$\text{or} \quad n_x = \frac{Lp_x}{2\pi}, \quad n_y = \frac{Lp_y}{2\pi}, \quad n_z = \frac{Lp_z}{2\pi}$$

¹¹With $\hbar = 1$ and $c = 1$.

Each (p_x, p_y, p_z) momentum state resides in the elemental volume in momentum space, $(2\pi/L)^3 = (2\pi)^3/V$, so that

$$\text{total phase space} = \frac{(2\pi)^3}{V} N_1$$

where N_1 is the number of momentum states available to one particle. Normalizing to one particle per *spatial* elemental volume and rewriting ‘total phase space’ as the integral over all momenta, we have

$$N_1 = \frac{1}{(2\pi)^3/V} \int dp_x dp_y dp_z / V = \frac{1}{(2\pi)^3} \int d^3\mathbf{p}$$

Next, we scale up to n particles:

$$N_{n-1} = \frac{1}{(2\pi)^{3(n-1)}} \int \prod_{j=1}^{n-1} d^3\mathbf{p}_j$$

from which¹² the density of states, the number of states per unit energy, is

$$\rho(E) = \frac{dN_{n-1}}{dE} = \frac{1}{(2\pi)^{3(n-1)}} \frac{d}{dE} \int \prod_{j=1}^{n-1} d^3\mathbf{p}_j \quad (2.33)$$

¹²As total momentum conservation constrains the n th particle, the number of available states is that of $n - 1$ free particles.

However, this is not satisfactory for high-energy physics, since we need to take into account the Lorentz contraction of the volume element in the usual NRQM wavefunction normalization.

dLips

The problem is solved by changing the quantum state normalization. Instead of normalizing to one particle per unit (spatial) volume $\int |\phi|^2 dV = 1$, we use

$$\int |\phi|^2 dV = 2E, \quad \text{for a particle of energy } E$$

For a particle of mass m with 4-momentum $p = (E, \mathbf{p})$ and spin (or helicity) λ , this corresponds to a momentum-space normalization of

$$\langle p', \lambda' | p, \lambda \rangle = 2E \delta_{\lambda\lambda'} \delta^3(\mathbf{p} - \mathbf{p}')$$

A useful identity, which shows the manifest Lorentz covariance of this choice, is

$$\frac{d^3\mathbf{p}}{2E} = \theta(E) \delta(p^2 - m^2) d^4p$$

Before giving the expressions for decay rates and cross sections, there is a somewhat tricky mathematical difficulty to be handled. In eqn 2.30,

we have a $\delta^4(\cdot)$ from overall 4-momentum conservation, which will be squared in the calculation. Formally, this is dealt with by using the identity

$$(2\pi)^4 \delta^4(p_f - p_i) = \int e^{ix(p_f - p_i)} d^4x \quad (2.34)$$

to replace one of the $\delta^4(\cdot)$ functions and then using $p_f = p_i$ plus the other $\delta^4(\cdot)$ to give $\int d^4x = VT$. These VT factors cancel with those that occur in the use of Fermi's Golden Rule, in which the normalized transition rate per unit volume and unit time appears: $w_{fi} = |S_{fi}|^2/VT$.

2.5.2 Decay rate

The partial decay rate of a state of mass M , at rest, into n particles is related to the reduced matrix element \mathcal{M} by

$$d\Gamma(M : m_1, \dots, m_n) = \frac{(2\pi)^4}{2M} |\mathcal{M}|^2 d\text{Lips}(P; p_1, \dots, p_n) \quad (2.35)$$

where $d\text{Lips}$ is the n -body phase space given by

$$d\text{Lips}(P; p_1, \dots, p_n) = \delta^4\left(P - \sum_{i=1}^n p_i\right) \prod_{i=1}^n \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} \quad (2.36)$$

Two-body decay rate

The simplest case is a two-body decay $X \rightarrow a + b$, where the masses of X, a , and b are M, m_a , and m_b , respectively. As the final-state integral is Lorentz-invariant, we are free to choose any frame. The rest frame of M is most convenient:¹³ $P_X = (M, \mathbf{0}), p_a = (E_a, \mathbf{p}_a), p_b = (E_b, \mathbf{p}_b)$. We have

$$\begin{aligned} \Gamma_{fi} &= \frac{(2\pi)^4}{2M} \int |\mathcal{M}_{fi}|^2 \frac{d^3\mathbf{p}_a}{(2\pi)^3 (2E_a)} \frac{d^3\mathbf{p}_b}{(2\pi)^3 (2E_b)} \\ &\quad \times \delta^3(\mathbf{p}_a + \mathbf{p}_b) \delta(M - E_a - E_b) \end{aligned}$$

Gathering constants and using the $\delta^3(\cdot)$ to remove one of the 3-momentum integrals, we obtain

$$\Gamma_{fi} = \frac{1}{32\pi^2 M} \int |\mathcal{M}_{fi}|^2 \frac{d^3\mathbf{p}_a}{E_a E_b} \delta(M - E_a - E_b)$$

In the CMS, we take $\mathbf{p}_a = \mathbf{k}$ and $\mathbf{p}_b = -\mathbf{k}$ and, using polar coordinates, we have

$$d^3\mathbf{p}_a \mapsto k^2 dk \sin\theta d\theta d\phi \mapsto k^2 dk d\Omega, \quad \text{where } k = |\mathbf{k}|$$

This gives

$$\Gamma_{fi} = \frac{1}{32\pi^2 M} \int |\mathcal{M}_{fi}|^2 \frac{k^2 dk d\Omega}{E_a E_b} \delta(M - E_a - E_b) \quad (2.37)$$

¹³It is the CMS frame of the decay products.

Since $\mathbf{p}_b = -\mathbf{p}_a = \mathbf{k}$, $E_b^2 = m_b^2 + k^2$ and eqn 2.37 becomes

$$\Gamma_{\text{fi}} = \frac{1}{32\pi^2 M} \int |\mathcal{M}_{\text{fi}}|^2 g(k) \delta(f(k)) dk d\Omega$$

where

$$g(k) = \frac{k^2}{E_a E_b}$$

$$f(k) = M_i - \sqrt{m_a^2 + k^2} - \sqrt{m_b^2 + k^2}$$

Denoting by p^* the value of $k = |\mathbf{p}_a|$ that conserves momentum, $g(k)$ can be integrated to give¹⁴

¹⁴Using the relationship

$$\delta(f(x)) = \left(\left| \frac{df}{dx} \right|_a \right)^{-1} \delta(x - a)$$

$$\begin{aligned} \int g(k) \delta(f(k)) dk &= \left(\left| \frac{df}{dk} \right|_{p^*} \right)^{-1} \underbrace{\int g(k) \delta(k - p^*) dk}_{g(p^*)} \\ &= \left(\left| \frac{df}{dk} \right|_{p^*} \right)^{-1} \frac{(p^*)^2}{E_a E_b} \end{aligned}$$

The inverted modulus term is obtained as follows:

$$\begin{aligned} \frac{df}{dk} &= -\frac{k}{\sqrt{m_a^2 + k^2}} - \frac{k}{\sqrt{m_b^2 + k^2}} = -\frac{k}{E_a} - \frac{k}{E_b} \\ &= -k \frac{E_a + E_b}{E_a E_b} \end{aligned}$$

so

$$\left(\left| \frac{df}{dk} \right|_{p^*} \right)^{-1} = \frac{1}{p^*} \frac{E_a E_b}{E_a + E_b}$$

We then have

$$\Gamma_{\text{fi}} = \frac{1}{32\pi^2 M} \frac{E_a E_b}{p^* (E_a + E_b)} \frac{(p^*)^2}{E_a E_b} \int |\mathcal{M}_{\text{fi}}|^2 d\Omega$$

By energy conservation, $E_a + E_b = M$, so, finally,

$$\Gamma_{\text{fi}} = \frac{p^*}{32\pi^2 M^2} \int |\mathcal{M}_{\text{fi}}|^2 d\Omega, \quad (2.38)$$

where the angular integral is over the solid angle of particle a .

Two-body decay kinematics

First, we calculate p^* from the original condition on the δ -function, $f(k) = 0$:

$$M = \sqrt{m_a^2 + p^{*2}} + \sqrt{m_b^2 + p^{*2}}.$$

¹⁵Try Exercise 2.5 at the end of the chapter for an alternative derivation of these results.

Some straightforward algebra yields¹⁵

$$p^* = \frac{1}{2M} \sqrt{[M^2 - (m_a + m_b)^2][M^2 - (m_a - m_b)^2]}. \quad (2.39)$$

Then, using $E^2 = p^2 + m^2$, the energies E_a and E_b are given by

$$E_a = \frac{M^2 + m_a^2 - m_b^2}{2M}, \quad E_b = \frac{M^2 + m_b^2 - m_a^2}{2M} \quad (2.40)$$

Note that energy and momentum conservation fix both the momenta and energies of the particles in a two-body decay. This is most easily seen in the rest frame of the decaying particle (or equivalently in the CMS frame of the decay products)—the only freedom is an azimuthal rotation about the common axis of the momenta of the two decay products.

Three-body decays

For a three-body decay, there are no longer enough constraints from 4-momentum conservation to determine the energies of the decay products, even in the rest frame of the parent particle. However, there are limitations that can be understood most easily by considering one decay product at rest with the other two particles then sharing the available energy as in a two-body decay. The most elegant approach to three-body decays is the Dalitz plot—a two-dimensional plot of either two decay particle energies or two decay particle invariant mass pairs.¹⁶ We shall use the latter, since the result is then manifestly Lorentz-invariant.

¹⁶The Dalitz plot is named after R. H. Dalitz, who invented it for the study of $K^0 \rightarrow 3\pi$ decays.

- Write the Lorentz sum of m_1 and m_2 , M_{12} , explicitly in terms of E_3 and p_3 :

$$\begin{aligned} M_{12}^2 &= (E_1 + E_2)^2 - (p_1 + p_2)^2 \\ &= (M - E_3)^2 - p_3^2 \end{aligned}$$

- Identify m_3^2 :

$$\begin{aligned} M_{12}^2 &= M^2 - 2ME_3 + E_3^2 - p_3^2 \\ &= M^2 + m_3^2 - 2ME_3 \end{aligned} \quad (2.41)$$

- Differentiate with respect to E_3 :

$$d(M_{12}^2) \propto dE_3 \quad (2.42)$$

- By similar reasoning,

$$d(M_{23}^2) \propto dE_1$$

Next, consider the infinitesimal three-body phase-space element $d\rho_3$, ignoring constant factors:

- As in the two-body case, one of the integrals over $d^3\mathbf{p}_i$ is removed using the 3-momentum $\delta^3(\cdot)$ function:

$$d\rho_3 \propto \frac{d^3\mathbf{p}_1 d^3\mathbf{p}_2}{E_1 E_2 E_3} \delta(E_1 + E_2 + E_3 - M)$$

- Change to spherical coordinates:

$$d^3\mathbf{p}_1 d^3\mathbf{p}_2 = dp_1 p_1 d\theta_1 p_1 \sin\theta_1 d\phi_1 dp_2 p_2 d\theta_2 p_2 \sin\theta_2 d\phi_2$$

- Redefine the solid angle:

$$\begin{aligned} d\Omega_1 d\Omega_2 &= \sin\theta_1 d\theta_1 \sin\theta_2 d\theta_2 d\phi_1 d\phi_2 \\ &= \sin\theta_1 d\theta_1 \sin\theta_{12} d\theta_{12} d\phi_1 d\phi_2 \end{aligned}$$

Then

$$d^3p_1 d^3p_2 = p_1^2 p_2^2 dp_1 dp_2 \sin\theta_1 d\theta_1 \sin\theta_{12} d\theta_{12} d\phi_1 d\phi_2$$

Simplify, noting that one part of the integration is trivial:

$$\int d^3\mathbf{p}_1 d^3\mathbf{p}_2 = 8\pi^2 \int p_1^2 p_2^2 dp_1 dp_2 \sin\theta_{12} d\theta_{12}$$

- Consider the momentum-squared of the back-to-back systems, \mathbf{p}_3 versus $\mathbf{p}_1 + \mathbf{p}_2$:

$$\mathbf{p}_3^2 = (\mathbf{p}_1 + \mathbf{p}_2)^2 = p_1^2 + p_2^2 + 2p_1 p_2 \cos\theta_{12}$$

- For a given $|\mathbf{p}_1|$ and $|\mathbf{p}_2|$, p_3 depends only on θ_{12} :

$$\begin{aligned} 2p_3 dp_3 &= -2p_1 p_2 \sin\theta_{12} d\theta_{12} \\ \sin\theta_{12} d\theta_{12} &\propto \frac{p_3}{p_1 p_2} dp_3 \end{aligned}$$

Put this result into $d\rho_3$, giving

$$\begin{aligned} d\rho_3 &\propto \frac{1}{E_1 E_2 E_3} p_1^2 p_2^2 dp_1 dp_2 \frac{p_3}{p_1 p_2} dp_3 \delta(E_1 + E_2 + E_3 - M) \\ &\propto \frac{p_1 p_2 p_3}{E_1 E_2 E_3} dp_1 dp_2 dp_3 \delta(E_1 + E_2 + E_3 - M) \end{aligned}$$

- Next, change variables, $p_i dp_i = E_i dE_i$ and remove one term using the δ -function:

$$d\rho_3 \propto dE_1 \cancel{dE_2} dE_3 \delta(E_1 + E_2 + E_3 - M)$$

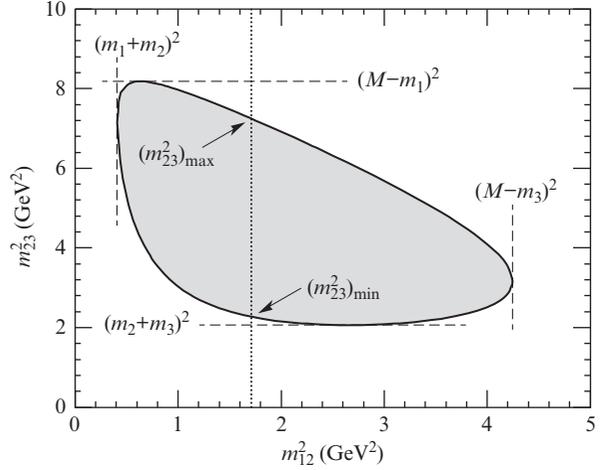


Fig. 2.3 A Dalitz plot [114].

- Finally, from eqn 2.42, $d(M_{12}^2) \propto dE_3$ and $d(M_{23}^2) \propto dE_1$:

$$d\rho_3 \propto d(M_{12}^2) d(M_{23}^2) \quad (2.43)$$

Figure 2.3 shows a Dalitz plot. The importance of the Dalitz plot is contained in this result—the phase space available in a three-body decay is uniform across the plot, here for M_{12}^2 versus M_{23}^2 . This means that dynamical structure, for example a resonance in one invariant mass pair, shows up as a region of higher or lower density in the Dalitz plot. These regions can then be seen clearly as peaks and troughs in appropriate one-dimensional projections. An example from the BaBar experiment [41] at the PEP-II e^+e^- storage ring is shown in Fig. 2.4. The Dalitz plot for

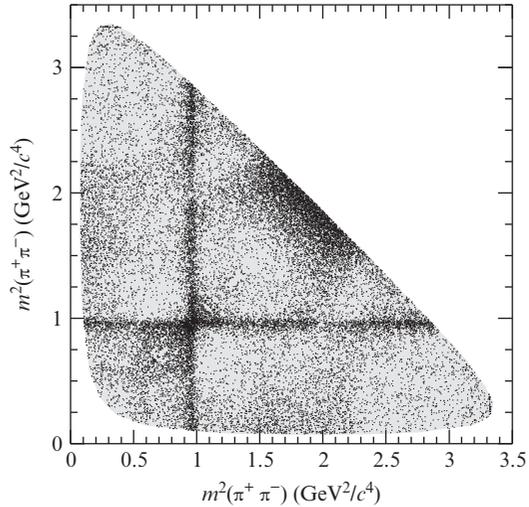


Fig. 2.4 $D_s^+ \rightarrow \pi^+ \pi^- \pi^+$ Dalitz plot from the BaBar experiment [41]. Resonant bands in the two $\pi^+ \pi^-$ invariant mass distributions can be seen clearly.

$D_s^+ \rightarrow \pi^+ \pi^- \pi^+$ is presented in terms of two π - π invariant mass combinations. The event distribution is clearly non-uniform, with two narrow $\pi\pi$ resonant bands at invariant masses corresponding to the $f_0(980)$ state. What might cause the accumulation of events around 1.9 GeV^2 on the diagonal?

2.5.3 Cross section

To calculate the total cross section for a collider process $a + b \rightarrow X$, we start from Fermi's golden rule:

$$\text{flux} \times \sigma(ab \rightarrow X) = \int w_{\text{fi}} \times [\text{final-state phase space}]$$

where $w_{\text{fi}} = |S_{\text{fi}}|^2/VT$ is the transition rate and VT is the total space-time volume. In a colliding-beam experiment, the initial-state particle flux is $2E_a 2E_b |\mathbf{v}_a - \mathbf{v}_b|$, where \mathbf{v}_i are the particle velocities.¹⁷

Inserting the expression from eqn 2.30 for S_{fi} in terms of the reduced matrix element into the above equation gives rise to the square of the overall 4-momentum conservation delta function. Formally, this is handled in the same way as for the decay rate calculation (eqn 2.34). The identity

$$(2\pi)^4 \delta^4(p_f - p_i) = \int e^{ix(p_f - p_i)} d^4x$$

is used to replace one of the $\delta^4(\cdot)$; then performing the integration with $p_f = p_{\text{fi}}$ on account of the other $\delta^4(\cdot)$ gives $\int d^4x = VT$. The VT factors then cancel to give

$$\sigma(ab \rightarrow X) = \frac{1}{2E_a 2E_b |\mathbf{v}_a - \mathbf{v}_b|} \int \text{dLips}(s : X) |\mathcal{M}(X : ab)|^2$$

where dLips is the Lorentz-invariant phase space, which for a final state with n_X particles is

$$\text{dLips}(s : X) = (2\pi)^4 \delta^4(p_X - p_a - p_b) \prod_{i=1}^{n_X} \frac{d^3\mathbf{k}_i}{(2\pi)^3 2k_i^0},$$

where $s = (p_a + p_b)^2$ and $p_X = \sum_i k_i$. For a total cross section to all final states X , an additional \sum_X is performed. For the unpolarized cross section for particles with spin, final spin states are summed and initial states averaged. This gives an additional term, $1/[(2S_a + 1)(2S_b + 1)]$, on the right-hand side, where S_a and S_b are the spins of particles a and b . If a differential cross section is required, then the relevant variables are excluded from the phase-space integral.

¹⁷The energy factors in the flux definition are a consequence of the Lorentz-invariant normalization.

Flux factor

The flux factor $2E_a 2E_b |\mathbf{v}_a - \mathbf{v}_b|$ is a Lorentz invariant and may be written in a number of ways. In the fixed-target frame, with b at rest, it is $2E_a 2m_b |\mathbf{v}_a|$. The following relations also hold:

$$2E_a 2E_b |\mathbf{v}_a - \mathbf{v}_b| = 4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = 2\sqrt{\lambda(s, m_a^2, m_b^2)}$$

where

$$\lambda(s, m_a^2, m_b^2) = (s - m_a^2 - m_b^2)^2 - 4m_a^2 m_b^2$$

Two-body scattering

Consider the special case of two-body scattering $\mathbf{a} + \mathbf{b} \rightarrow \mathbf{a}' + \mathbf{b}'$ with 4-momenta $p_a = p_1$, $p_b = p_2$, $p_{a'} = p_3$, $p_{b'} = p_4$. The cross section in the fixed-target frame (particle b at rest) is given by

$$d\sigma = \frac{|\mathcal{M}|^2}{2E_1 2m_2 |\mathbf{v}_1|} d\text{Lips}(s; p_3, p_4) \quad (2.44)$$

where in this case

$$d\text{Lips}(s; p_3, p_4) = (2\pi)^4 \delta^4(p_3 + p_4 - p_1 - p_2) \frac{d^3\mathbf{p}_3}{2E_3(2\pi)^3} \frac{d^3\mathbf{p}_4}{2E_4(2\pi)^3}$$

with $s = (p_3 + p_4)^2 = (p_1 + p_2)^2$. $d\text{Lips}$ may be evaluated by using the $\delta^4(\cdot)$ to integrate over four of the six variables $d^3\mathbf{p}_3 d^3\mathbf{p}_4$, and doing this in the CMS frame gives¹⁸

$$\frac{d\sigma}{d\Omega} = \frac{1}{(8\pi w)^2} |\mathcal{M}|^2 \frac{p'}{p} \quad (2.45)$$

where $w = \sqrt{s}$ is the total CMS energy and p and p' are the magnitudes of the initial and final CMS 3-momenta, respectively. The above cross section has dimensions of $[\text{energy}]^{-2}$. To get back to physical units of area, one must multiply by $(\hbar c)^2 = 0.389 \text{ mb GeV}^2$.

2.5.4 Breit–Wigner

In nuclear physics, τ might be anywhere from 10^{-3} to 10^{+32} s but particle physics typically deals with short timescales: 10^{-23} – 10^{-8} s. As we have seen above, the lifetime of an unstable state is inversely related to the total decay rate Γ ($\Gamma\tau \sim 1$). This is a form of the time–energy uncertainty relation: the uncertainty in lifetime translates to an uncertainty in mass. The state is said to have a **natural width**

$$\Gamma = 1/\tau \quad \text{and} \quad N(t) \propto e^{-\Gamma t} \quad (2.46)$$

For long-lived particles (meaning $\sim 10^{-13}$ s or more), the natural width is so small that it is better to quote mass and lifetime.

¹⁸Exercise 2.9 covers this calculation.

A simple model for such states gives rise to the Breit–Wigner line shape (Fig. 2.5). We proceed as follows:

- The wavefunction of a state of energy E_R and lifetime $\tau = 1/\Gamma$ is

$$\begin{aligned}\psi(t) &= \psi(0)e^{-iE_R t}e^{-t/2\tau} \\ &= \psi(0)e^{-t(iE_R + \Gamma/2)}\end{aligned}$$

- The intensity $\psi\psi^*$ follows the exponential decay law $I \propto e^{-\Gamma t}$. The amplitude as a function of E is derived from the Fourier transform:

$$\begin{aligned}\chi(E) &= \int \psi(t)e^{iEt} dt \\ &= \psi(0) \int e^{-t[(\Gamma/2) + i(E_R - E)]} dt \\ &\propto \frac{1}{(E - E_R) - i\Gamma/2}\end{aligned}$$

- The cross section $\sigma(E) \propto \chi\chi^*$:

$$\sigma(E) = \sigma_{\max} \frac{\Gamma^2/4}{(E - E_R)^2 + \Gamma^2/4} \quad (2.47)$$

The value of σ_{\max} can be found using a heuristic argument from wave optics following Perkins [117]. The angular momentum of a particle with momentum p about a scattering centre may be written as $L = pb$, where b is the impact parameter.¹⁹ Particles in a beam of fixed momentum will have b in the range $(0, L/p)$ and will impact on the target plane within a circular disc of area πb^2 . The wavefunction of the incident particles can be expanded as a sum of *partial waves* with quantized angular momentum $L = l$, where l is an integer. The fraction of the beam with $L \in (l, l+1)$ hits the target plane within an annulus of area $\pi[(l+1)^2 - l^2]/p^2 = (2l+1)\pi(1/p)^2$. To take account of elastic and totally absorptive scattering, the elastic scattering amplitude is doubled, leading to a factor of 4 in this expression.²⁰ For scattering through the l th partial wave, the result is $\sigma_{\max} = 4\pi(1/p)^2(2l+1)$.

A few more details are needed:

- So far, the spin of the particles has been ignored. Since the Breit–Wigner formula is a total cross section, we sum over the final-state spins and average over the initial-state spins, giving a factor

$$\frac{2J+1}{(2S_a+1)(2S_b+1)}$$

For scattering through a resonant state with spin J , $l \rightarrow J$.

- The expression in eqn 2.47 for the energy dependence is not in a Lorentz-invariant form. This is rectified by multiplying it by $(E + E_R)^2$ and using the approximation $E \approx E_R = M_0$ where appropriate.

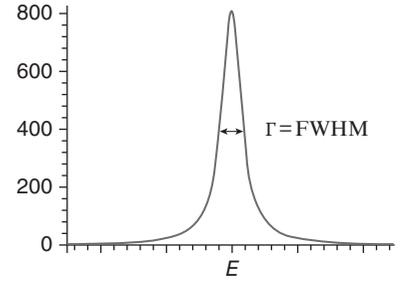


Fig. 2.5 Breit–Wigner line shape.

¹⁹The impact parameter is defined as the perpendicular distance from the scattering centre to the direction of travel of the object.

²⁰For more details on partial wave analysis, see the Further Reading at the end of the chapter.

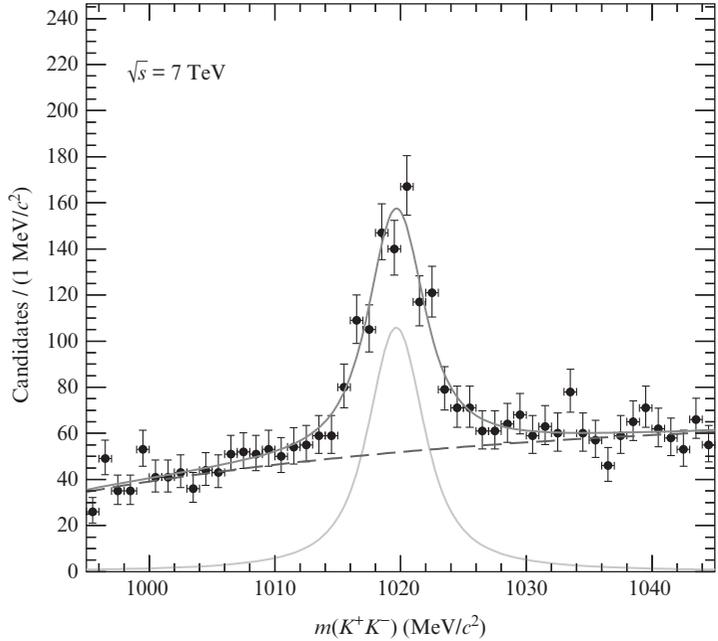


Fig. 2.6 Breit–Wigner line shapes for $\phi \rightarrow K^+K^-$, from the LHCb experiment at the LHC [100]. The Breit–Wigner line shape is superimposed on a gradually rising background shown by the dashed line.

- Often, a resonance will occur in many different scattering processes (or channels), and to allow for this, the expression in eqn 2.47 is further multiplied by the branching ratios BR_{in} and BR_{out} for the entrance and exit channels.

The result is

$$\sigma(s) = \sigma_{\text{max}} \frac{M_0^2 \Gamma^2}{(s - M_0^2)^2 + M_0^2 \Gamma^2} \quad (2.48)$$

where

$$\sigma_{\text{max}} = \frac{4\pi (2J + 1) \text{BR}_{\text{in}} \text{BR}_{\text{out}}}{p^2 (2S_a + 1)(2S_b + 1)} \quad (2.49)$$

Two examples of Breit–Wigner line shapes are shown in Figs. 2.6 and 2.7, for the decays $\phi \rightarrow K^+K^-$ and $K^* \rightarrow K^+\pi^-$, respectively. Both states have comparable masses ($\phi \sim 1020 \text{ MeV}$, $K^* \sim 900 \text{ MeV}$), but the K^* width²¹ is about 50 MeV, while that of the ϕ is only 4.3 MeV. Try Exercise 2.8 to help understand this.

²¹The full width at half the maximum height of the peak (FWHM).

2.6 Luminosity and event rates

In Section 2.5.3, a relationship among event rate, flux, and cross section was introduced and elaborated, particularly for two-body scattering. The strategies for optimizing and measuring the luminosity in colliders are

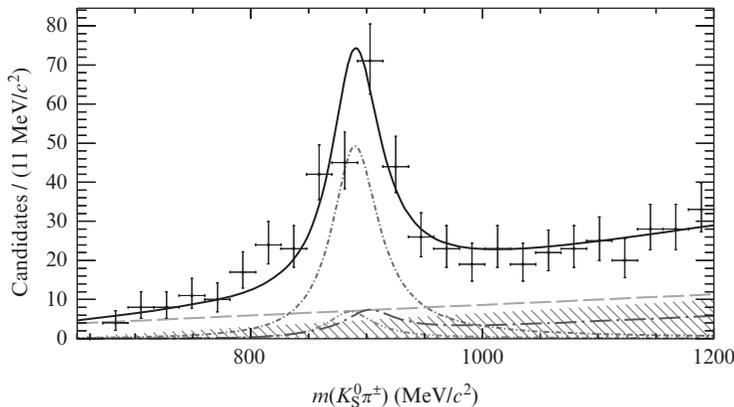


Fig. 2.7 Breit–Wigner line shapes for $K^* \rightarrow K^+\pi^-$, from the LHCb experiment at the LHC [105]. The overall background under the Breit–Wigner line shape is shown by the dashed line.

described in Section 3.4. The critical issue of ‘triggering’ (i.e. selecting the interesting events to keep) is introduced in Section 4.10 and discussed in slightly more detail for the LHC in Section 13.2.

2.7 Group theory

This section provides a brief introduction to group theory, with emphasis on the groups of relevance to physics and particularly particle physics. The mathematical definition of a group G considers a set of objects (a, b, c, \dots) with a multiplication rule satisfying the following:

- for objects $a, b \in G$, the product ab is also a member of G ;
- it is associative: $a(bc) = (ab)c$ for $(a, b, c) \in G$;
- G contains a unit element e such that $ae = ea = a$ for all $a \in G$;
- for all $a \in G$, there exists an inverse $a^{-1} \in G$ satisfying $aa^{-1} = a^{-1}a = e$.

There are many different types of group of relevance to physics:

S_n the group of permutations of n objects ($n!$ elements), which involves discrete operations;

T_2 translations of two-dimensional vectors in a plane, $T_{\mathbf{a}} : \mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} + \mathbf{a}$;

R_3 rotations of three-dimensional vectors, $x'_i = R_{ij}x_j$. R_3 conserves distance, which requires: $\det(R) = 1$ and $\sum_j R_{ij}R_{jk} = \delta_{ik}$ (R is a proper orthogonal matrix);

$U(n)$, $SU(n)$ the groups of unitary and unimodular $n \times n$ matrices U satisfying $UU^\dagger = U^\dagger U = I_n$, with $\det(U) = e^{i\phi}$ for $U(n)$ and $\det(U) = +1$ for $SU(n)$.

²²For T_2 , a two-dimensional vector translation in a plane requires the distance that the centre of the vector is moved in the plane and the angle of rotation (also within the plane) with respect to the original direction of the line. For R_3 , a rotation in three dimensions requires the direction of the axis about which the rotation takes place (two direction cosines) and the angle of rotation about this axis.

T_2 and R_3 depend on two and three parameters, respectively.²² These numbers are known as the *order* of the group. For T_2 , the sequence in which two successive translations is applied does not affect the outcome: $T_2(\mathbf{a})T_2(\mathbf{a}') = T_2(\mathbf{a}')T_2(\mathbf{a})$; such a group is called *Abelian*. The groups R_3 and $SU(n)$ are non-Abelian, so the sequence of successive operations matters.

2.7.1 Lie groups

Consider a matrix group $\{A\}$. An element A must have an inverse, and hence $\det(A) \neq 0$; thus, there exists a matrix α such that

$$A = \exp(\alpha) = 1 + \alpha + \alpha^2/2 + \dots$$

and α is the logarithm of A . The set of all matrices α whose exponentials belong to a group G is known as the *Lie algebra* of G . Using the definition of the exponential series, we have

$$A = \lim_{k \rightarrow \infty} \left(I + \frac{\alpha}{k} \right)^k \quad \text{and inverse} \quad \alpha = \lim_{k \rightarrow \infty} k(A^{1/k} - I)$$

For large k , the matrix $I + \alpha/k$ is an operator of the group and gives A by iteration. It is an *infinitesimal operator*, since for large k it differs from the identity operator by an infinitesimal amount.

The product of two elements of a group is a member of the group, which for a finite group must then be expressible as a linear combination of the n generators of the Lie algebra.²³ This provides a set of relations among the commutators of the basis elements $\{g_i\}$, $i = 1, \dots, n$:

$$[g_i, g_j] = ic_{ij}^k g_k \quad (2.50)$$

There will be $n(n-1)/2$ such relations, and the c_{ij}^k are known as the *structure constants* of the group.

2.7.2 $U(n)$ and $SU(n)$

The Lie groups $U(n)$ and $SU(n)$ occur in a number of different contexts in particle physics. $U(1)$ and $SU(2)$ are used in the construction of the electroweak sector of the standard model. $SU(2)$ and $SU(3)$ occur as approximate symmetries used to classify nuclear and particle states. The strong force of QCD is based on an exact $SU(3)$ of eight gluons providing the strong force binding the quarks into hadrons.

²⁴That is, $UU^\dagger = U^\dagger U = I$.

$U(n)$ The set of all $n \times n$ unitary²⁴ matrices form the $U(n)$ group. Because $UU^\dagger = I$, $\det(UU^\dagger) = 1$, and so we have

$$\begin{aligned} \det(UU^\dagger) &= \det(U) \det(U^\dagger) \\ &= (\det U)(\det U)^* \\ &= |\det U|^2 = 1 \end{aligned}$$

$$\text{hence} \quad \det U = e^{i\phi}$$

$SU(n)$ This is the subset of $U(n)$ for which $\phi = 0$, i.e. $\det(U) = 1$. Identifying $U = e^{ig}$, where g is the *generator* of the operator U , and using the relation

$$\det(e^{ig}) = e^{i\text{Tr}(g)}$$

it follows that e^{ig} is an element of $SU(n)$ if, and only if, g is traceless: $\text{Tr}(g) = 0 \Rightarrow e^{i\text{Tr}(g)} = 1$.

Once the relevant group has been identified, properties of a model can be inferred from the properties of that group:

- for $SU(n)$, the number of independent generators is $n^2 - 1$;
- so for $SU(2)$ this is $2^2 - 1 = 3$, and for $SU(3)$ it is $3^2 - 1 = 8$.

2.7.3 $SU(2)$

The simplest $SU(n)$ group is $SU(2)$. It is the group of 2×2 unitary matrices U with $\det(U) = 1$. Writing $U = e^{ih}$, unitarity gives

$$UU^\dagger = e^{ih}e^{-ih^\dagger}.$$

From the unitarity condition, $(UU^\dagger)^\dagger = U^\dagger U = I$, so U and U^\dagger commute and it follows that h and h^\dagger also commute. Under these conditions,

$$e^{ih}e^{-ih^\dagger} = e^{i(h-h^\dagger)} = I = e^0$$

so $h - h^\dagger = 0$. These are the conditions for h to be a Hermitian matrix.

$SU(2)$ has three generators, and the three 2×2 traceless Hermitian matrices making up the Lie algebra are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.51)$$

Any 2×2 traceless Hermitian matrix h can be written as a linear combination of the three Pauli matrices with real parameters λ_i :

$$h = \lambda_1\sigma_1 + \lambda_2\sigma_2 + \lambda_3\sigma_3 \quad (2.52)$$

The commutation relations of the Pauli matrices are

$$[\sigma_1, \sigma_2] = 2i\sigma_3, \quad [\sigma_2, \sigma_3] = 2i\sigma_1, \quad [\sigma_3, \sigma_1] = 2i\sigma_2 \quad (2.53)$$

So for this group the structure constants are all 2. The angular momentum vector of a spin- $\frac{1}{2}$ particle is $J_i = \frac{1}{2}\sigma_i$, with the usual angular momentum commutation relations $[J_i, J_j] = i\epsilon^{ijk}J_k$.

The Pauli matrices provide the *fundamental* or *irreducible* matrix representation of the group $SU(2)$. From these, representations of higher angular momentum states can be constructed—see the Further Reading.

2.7.4 Combining states

²⁵A familiar example is how to combine angular momentum states (representations of the three-dimensional rotation group R_3).

Representation theory provides the rules for how to combine states.²⁵ Some examples of the arithmetic for combining $SU(n)$ states are as follows:

- $SU(2)$:

$$2 \otimes 2 = 4 = 3 \oplus 1$$

(3 symmetric states, 1 antisymmetric state);

$$2 \otimes 2 \otimes 2 = 8 = 4 \oplus 2 \oplus 2$$

(4 symmetric states, 2 mixed-symmetric and 2 mixed-antisymmetric).

- $SU(3)$:

$$3 \otimes 3 \otimes 3 = 27 = 10 \oplus 8 \oplus 8 \oplus 1$$

(10 symmetric states, 8 mixed-symmetric, 8 mixed-antisymmetric and one antisymmetric state).

Further details of $SU(2)$ and $SU(3)$ and their use in particle physics are given in Chapter 5. See also the exercises at the end of this chapter.

Chapter summary

- Symmetries and invariance, addition of angular momentum.
- Clebsch–Gordan coefficients and branching ratios.
- Lorentz-invariant variables, cross section, and decay rate.
- Two- and three-body kinematics, Dalitz plots.
- Unstable states and Breit–Wigner resonance, event rates.
- Group theory, the groups $U(1)$, $SU(2)$, $SU(3)$.

Further reading

- Gasiorowicz, S. (1974). *Quantum Physics*. Wiley.
- Muirhead, H. (1974). *The Special Theory of Relativity*. Macmillan.
- Halzen, F. and Martin, A. (1974). *Quarks and Leptons*. Wiley. Gives a good description of partial waves.

Exercises

- (2.1) What is the minimum beam energy needed to produce a W boson in a fixed-target collision? Compare this value with the equivalent beam energy of a symmetric $\bar{p}p$ collider.
- (2.2) Consider the production of neutrinos from decays of π^\pm . The π^\pm are produced by collisions of protons of energy E on a stationary target. Estimate the minimum energy required to produce neutrinos of energy 10 GeV.
- (2.3) The ρ^0 decays strongly to $\pi^+\pi^-$, but not to $\pi^0\pi^0$. Why?
- (2.4) Starting from eqn 2.15, determine the elements of the rotation matrix for spin- $\frac{1}{2}$ particles.
- (2.5) Consider two-body decay $A \rightarrow B + C$. In the rest frame of A , use energy conservation ($M_A = E_B + E_C$) and 3-momentum conservation (decay products B and C must travel in opposite directions) to give an alternative derivation of the expressions for E_A and E_B in eqn. 2.40. Finally, derive eqn. 2.39 for p^* (magnitude of B or C 3-momentum).
- (2.6) Starting from the definitions of s and t , eqn. 2.23, derive eqns 2.24 and 2.25 for the case of πp scattering in a fixed-target geometry.
- (2.7) Referring to Fig. 2.4, the Dalitz plot for the decay $D_s^+ \rightarrow \pi^+\pi^-\pi^+$, can you explain the accumulation of events around a mass-squared of 1.9 GeV^2 on the diagonal? The PDG data tables may be useful.
- (2.8) Why does the $\phi(1020)$ have such a small width and what might this imply about its quark content?
- (2.9) Work through the calculation to get from eqn 2.44 to the expression in eqn 2.45 for $d\sigma/d\Omega$. Think carefully about how to use the overall $\delta^4(\cdot)$ constraint from 4-momentum conservation.
- (2.10) Using the definitions of $U(n)$ and $SU(n)$, show that the unitary matrices used to represent $SU(n)$ require $n^2 - 1$ real numbers.
- (2.11) Show that for a Lie group of order n , there will be $n(n - 1)/2$ commutation relations defining the structure constants.

3

Accelerators

3.1 Radiofrequency acceleration	48
3.2 Beam optics	54
3.3 LHC dipole magnets	59
3.4 Colliders and fixed-target accelerators	62
3.5 Luminosity	63
3.6 $\bar{p}p$ colliders	65
3.7 Future accelerators	67
Chapter summary	68
Further reading	68
Exercises	69

¹An essential step in electronic chip fabrication.

Accelerators are devices that accelerate charged particles to a broad range of energies from keV to TeV. This chapter is a very brief introduction to accelerators in particle physics and high-energy nuclear physics. There are only a few accelerators used for particle or nuclear physics, but about 30 000 accelerators are currently used worldwide for other very important applications. There are about 9000 accelerators used in cancer therapy, 9500 in ion implantation,¹ 4500 for cutting and welding, 2000 for electron-beam and X-ray sources, 1000 for neutron generators, and more in other fields. Each type of accelerator is built using the technology most suitable for its particular application. In this very brief introduction, we will focus only on synchrotrons and linear accelerators, since these are the typical choices in particle physics. Some excellent introductory textbooks on accelerator physics are given in the Further Reading at the end of this chapter.

3.1 Radiofrequency acceleration

3.1.1 Electric and magnetic fields

Particles are accelerated by the electric field only,

$$\mathbf{E} = -\nabla\varphi - \frac{\partial\mathbf{A}}{\partial t} \quad (3.1)$$

where φ is the scalar potential and \mathbf{A} the vector potential. In particle physics accelerators, the time-dependent vector potential \mathbf{A} is the source of the accelerating field \mathbf{E} . The simplest realization is a cylindrical structure, sketched in Fig. 3.1, and is called a pill-box cavity. Microwave radiation (with frequency in the MHz–GHz range) produced in a device called a klystron (see Section 3.1.6) is guided to the pill-box cavity, where it forms a standing wave; i.e. the pill-box cavity acts as a resonator. In free space, electromagnetic waves can only have transverse electric and magnetic fields with respect to the direction of propagation; however, inside a cavity, we can have transverse electric (TE) or transverse magnetic (TM) modes, indicating fields that have either longitudinal magnetic or electric fields, respectively. Since we need a longitudinal electric field to accelerate charged particles, only the TM modes will be useful. The modes can be found by solving Maxwell's equations in free space without free charges or currents, subject to the usual boundary conditions at

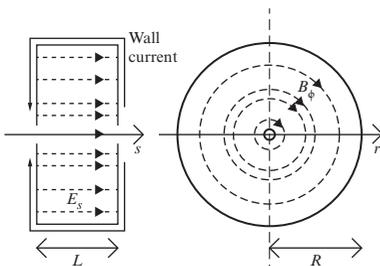


Fig. 3.1 A pill-box cavity.

the conducting surfaces of the cavity. These conditions ensure that the longitudinal component of the \mathbf{E} field and the perpendicular component of the \mathbf{B} field vanish at the surface of a conductor.

The useful (i.e. accelerating) modes of this cavity are TM_{nlm} , where the indices, n , l , and m refer to the field variations along the usual polar coordinates ϕ (azimuthal), r (radial), and s (longitudinal, i.e. along the beam direction). Since the radial variations are given by Bessel functions and we require a non-zero component of the electric field in the longitudinal direction s on-axis (i.e. $E_s(r=0) \neq 0$), and we must satisfy the boundary conditions at the surface of the conductor, we need the $l=1$ modes. We want to select the modes that minimize the energy stored (and hence the electricity costs) for a given accelerating gradient. This means that we use the TM_{010} mode. This mode has only two components, the electric field E_s in the direction of the acceleration (s) and the azimuthal component of the magnetic field B_ϕ in the cavity (as indicated in Fig. 3.1), which are oscillating with the radiofrequency (RF) frequency ω :

$$E_s \sim J_0(kr)e^{i\omega t}, \quad B_\phi \sim J_1(kr)e^{i\omega t} \quad (3.2)$$

where J_0 and J_1 are the lowest-order Bessel functions and $k = 2\pi/\lambda$ is the wavenumber. The requirement that $E_s(R) = 0$, where R is the radius of the pill box, determines the allowed value of k from the zeros of the Bessel function. The first zero of the J_0 Bessel function is $J_0(2.40) = 0$, therefore $kR = 2.40$ and hence $\lambda = (2\pi R)/2.4$. So for the $l=1$ mode, $\lambda \simeq 2.62R$. The amplitudes J_0 and J_1 depend on the radial coordinate r as sketched in Fig. 3.2.

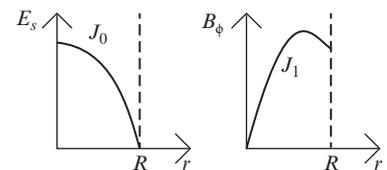


Fig. 3.2 Amplitudes of the TM_{010} fields in a pill-box cavity.



Fig. 3.3 The child accelerates the roundabout by only pushing at the correct phase.

3.1.2 Circular accelerators and synchronicity

Suppose now that we want to accelerate protons in this pill-box cavity along the direction s , which goes from left to right. We need to inject these protons into the cavity when the electric field component E_s can accelerate them, i.e. when it is positive. After a time $t = \pi/\omega$, the field changes sign and injected protons would be decelerated. If we further assume that the pill-box cavity is an accelerating structure of a synchrotron where protons go around on a closed trajectory, approximately a circle,² then we will end up with the synchronicity condition

$$\omega = N\omega_{\text{rev}}. \quad (3.3)$$

The RF frequency needs to be an integer (N) multiple of the revolution frequency ω_{rev} with which protons go around in the synchrotron. This is demonstrated in the cartoon in Fig. 3.3. The value of N is chosen for practical reasons—for example to make the RF frequency lie in the range where components such as amplifiers are available. This means that N is typically very large (e.g. at LEP, $N = 31\,320$). This defines the number of ‘buckets’ in which we can potentially store stable beams. However, we usually only want to inject particles into a much smaller number of buckets³. In a synchrotron, protons go through the accelerating cavities

²There will be straight sections, for example in the experimental halls where particle physics detectors are located.

³The groups of particles in filled buckets are called bunches. At the LHC, the spacing between buckets is 2.5 ns, but only $\sim 10\%$ of buckets are filled with protons.

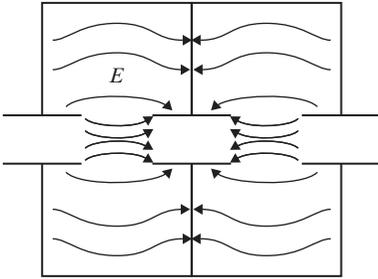


Fig. 3.4 An accelerating cavity consisting of two pill-box cavities. A proton bunch in the left-hand gap will be accelerated. A proton bunch in the middle drift tube will be shielded from the decelerating field. When it emerges into the right-hand gap, the field has changed sign.

many times, gaining energy at every passage. The magnetic field guiding them through the accelerator (see Section 3.2) changes along with the acceleration, keeping them on the same orbit.

One can combine many pill-box cavities, stacking them one after another as sketched in Fig. 3.4. Protons are then accelerated in gaps between the drift tubes (see Fig. 3.4) when the accelerating electric field points in the right direction and then ‘hide’ inside the drift tubes, isolated from the electric field when it points in the wrong direction, emerging in the next pill box when the electric field is again pointing in the right direction.

The fact that beam particles need to enter an accelerating cavity at the right time leads to the bunched structure of the beam. This is demonstrated in Fig. 3.5 in more detail. A proton, for example, with nominal momentum p and travelling along the nominal path is called a synchronous particle and its trajectory is the reference trajectory. It arrives at the entrance to the cavity at point M_1 . After going through the cavity, its energy is increased by ε (here we are assuming that $\omega L/c \ll 1$). Another proton arrives a little earlier at point P and its energy is increased by a smaller amount than ε , so when it enters the cavity again, after the next revolution, it will not be that early in comparison with the synchronous particle.⁴ Yet another proton arrives a little late at point P' . This proton will gain more energy than ε , so, after the next revolution, it will not be that late. In a synchrotron in which we combine the accelerating cavities with magnetic fields, these different energy gains and losses will lead to oscillations (‘synchrotron oscillations’).

In contrast, considering points Q and Q' in relation to point N_1 , one can see that a proton arriving early with respect to N_1 will gain more energy and one arriving late will gain less energy than the proton arriving at N_1 ; thus, every revolution, these protons will be further and further apart from each other, eventually escaping from the accelerator.⁵ So M_1 and M_2 are stable points where bunches of beam particles can be located and N_1 and N_2 are unstable points.

This simplified discussion of phase stability needs to be expanded to take into account competing changes: the speed of the particles and the

⁴For simplicity, we assume that the paths around the accelerator are of the same length for all particles considered. In reality, different momenta/energies lead to different paths, which also need to be taken into account.

⁵For electrons, we would need to take account of energy losses due to synchrotron radiation and differences among trajectories of particles with different momenta. The conclusion would be that N_1 and not M_1 would correspond to a stable reference trajectory for a bunch of electrons.

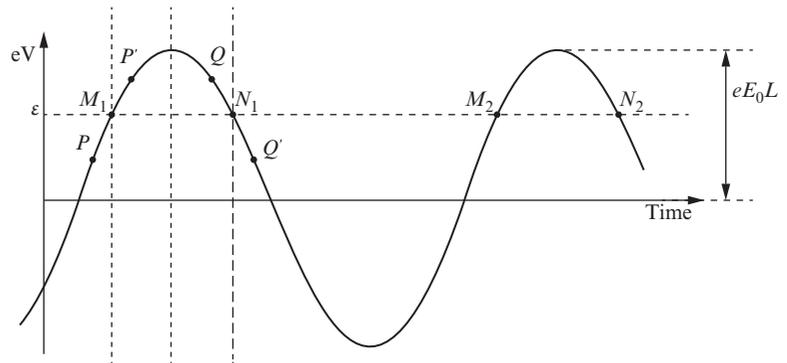


Fig. 3.5 Energy gain in a cavity as a function of the arrival time or relative phase of a beam particle with respect to the oscillating electric field in the cavity. E_0 is the accelerating electric field amplitude, L the cavity length, and e the charge of accelerated particle. Other symbols are explained in the text.

radius of the orbit. The frequency is simply related to the speed v and the radius R by

$$f = \frac{v}{2\pi R}. \quad (3.4)$$

Both v and R depend on the momentum of the particle. From relativistic kinematics, we know that

$$p = \frac{mv}{\sqrt{1-v^2}} \quad (3.5)$$

When the RF acceleration increases the momentum of the particle, it will also cause it to follow a slightly different orbit. This change ΔR in the radius is defined by the dispersion D , which for a certain change Δp in the momentum gives⁶

$$\Delta R = D \frac{\Delta p}{p} \quad (3.6)$$

⁶The value of the dispersion depends on the type and strength of the magnetic focusing. See Wilson in Further Reading for details.

It is conventional to define the ‘slip factor’, which is given by the fractional rate of change with frequency divided by the fractional change in momentum:

$$\eta_{\text{RF}} = \frac{\Delta f/f}{\Delta p/p} \quad (3.7)$$

Substituting from eqns 3.5 and 3.6, we can evaluate (see Exercise 3.2) the two terms in eqn 3.7:

$$\eta_{\text{RF}} = \frac{1}{\gamma^2} - \frac{D}{R_0} \quad (3.8)$$

where R_0 is the radius of the reference trajectory. Therefore, for injection at low momentum, for which $\gamma \sim 1$, for a typical proton synchrotron, $\eta_{\text{RF}} > 0$. However, as the momentum increases, we will reach a transition in which $\eta_{\text{RF}} = 0$ and, for higher momentum, $\eta_{\text{RF}} < 0$. This implies that the region of phase stability flips when we cross the transition defined by

$$\frac{1}{(\gamma_{\text{transition}})^2} = \frac{D}{R_0} \quad (3.9)$$

Therefore, we need to change the RF phase as we cross such transitions.⁷ The concept of phase stability discussed here is one of the key ideas that enabled the successful operation of high-energy accelerators.

⁷This can be done sufficiently quickly that the particle losses are negligible.

3.1.3 Accelerating-cavity design and Q factor

For efficient operation of an accelerating cavity,⁸ i.e. a standing-wave resonator, one requires that the energy be transferred from the resonator to the accelerated beam and not dissipated into the environment by losses to the walls of the cavity and by radiation to the environment. This means that the Q value, i.e. the quality factor defined as the ratio

⁸In this subsection we will work in SI units

of the average energy stored in the cavity, U , to the average energy dissipated in one oscillation period, U_d , is very large. The energy stored alternates between the electric and magnetic fields, but we can calculate this from the peak value of the magnetic field, \mathbf{H}_0 :

$$U = \frac{\mu_0}{2} \int |\mathbf{H}_0|^2 d^3\mathbf{r} \quad (3.10)$$

⁹This is a simplified calculation that neglects radiation losses.

We can calculate U_d from the Ohmic losses at the surface of the cavity.⁹ For a good conductor ($\sigma/\epsilon\omega \gg 1$), we can neglect the displacement current. Using Amperè's law, we can show that $|\mathbf{H}| = j_{\text{surf}}$, where j_{surf} is the surface current per unit length. The power dissipated, I^2R , can be evaluated as a surface integral

$$P = \frac{1}{2} \int |\mathbf{H}_0|^2 R_{\text{surf}} dS \quad (3.11)$$

where the surface resistance $R_{\text{surf}} = 1/\sigma\delta$ and the skin depth $\delta = \sqrt{2/\mu\sigma\omega}$. The energy dissipated over one period $T = 2\pi/\omega$ is then given by

$$\Delta W = \pi \sqrt{\frac{\mu}{2\sigma\omega}} \int |\mathbf{H}_0|^2 dS \quad (3.12)$$

¹⁰For superconducting cavities, we have to consider the effective surface resistance, so there are still losses; however these are orders of magnitude smaller for superconducting nickel compared with copper, which suffers from Ohmic losses.

Comparing eqns 3.10 and 3.12, we can see that to maximize the Q value, we need a high frequency, a high conductivity, and a large ratio of volume to surface area. The RF frequency used is usually in the range 100 MHz to ~ 10 GHz.¹⁰

Next we need to consider the fact that the electric field is varying while the particle crosses the cavity. The electric field for the $n = 0$ mode on axis ($r = 0$) depends on time as

$$E_s = E_0 \cos(\omega t) \quad (3.13)$$

For an ultrarelativistic particle crossing the cavity, the position along the axis of the cavity is simply $s = ct$ and its speed does not change, but it gains energy over the length of the gap, G :

$$\Delta W = \int_{-G/2}^{G/2} E_0 \cos(\omega s/c) ds = \frac{\sin(\omega G/2c)}{\omega G/2c} E_0 G \quad (3.14)$$

It is clear from eqn 3.14 that for efficient acceleration, we need the gap length to be significantly less than the wavelength. This helps keep the bunches in the accelerating phase and prevents slippage into the decelerating phase. On the other hand, we have seen that to obtain a high Q value (and hence minimize Ohmic losses), we need a large ratio of volume to surface area. The RF cavities that are used in high-energy accelerators have shapes optimized to meet these two requirements, which include a short gap length and a large volume. A (very schematic) sketch of a cross section of a cavity is shown in Fig. 3.6. Typically, many cavities are combined in one structure.

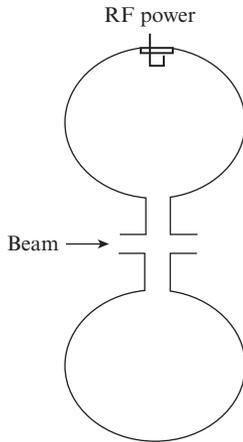


Fig. 3.6 Schematic cross-section of an RF cavity (not to scale). Note the small accelerating gap and the relatively large volume. The RF power enters through an insulating ceramic window and couples into the cavity.

Good-quality accelerating cavities that can be produced in large quantities have accelerating fields in the region of 20–30 MV m⁻¹. Single cavities might achieve up to 100 MV m⁻¹, which is a breakdown limit (or beyond) for most materials.¹¹ To achieve larger accelerating fields, one needs a different approach. Fields up to 100 GV m⁻¹ can be obtained in plasma (no walls to break down), where electrons can be displaced from quasistationary ions. This is a very active research area that eventually might lead to a new generation of accelerators.

3.1.4 Synchrotron radiation energy loss

When charged particles are accelerated in a circular machine, they lose energy by synchrotron radiation. For an ultrarelativistic particle of mass m and Lorentz factor γ , in an orbit with radius of curvature ρ , the power emitted in the form of synchrotron radiation is

$$P = \frac{2}{3} \frac{r_c m \gamma^4}{\rho^2} \quad (3.15)$$

where $r_c = e^2/(4\pi\epsilon_0 mc^2)$ is the classical radius of the particle. As the synchrotron radiation scales as γ^4 , it will generally be negligible for protons, but the losses for electron machines will be very significant. The energy loss grows as the fourth power of the energy, and therefore there is a limit to the energy reach of circular electron machines. Although synchrotron radiation can be reduced by increasing the radius of the machine, this becomes prohibitively expensive at high energies. LEP is generally considered to be the highest-energy circular electron accelerator that will be built: higher-energy electron–positron machines will be linear colliders, for which the synchrotron radiation is negligible.¹² While synchrotron radiation is a problem for particle physics applications, it turns out to have many uses in other areas of science. The synchrotron radiation in the laboratory frame is forward-peaked around the electron direction and provides a very high-brightness X-ray source. Dedicated electron rings are built with ‘wiggler’ magnets to increase the synchrotron radiation. The X-rays are used in condensed matter physics, biology, medical applications, and other fields.

3.1.5 Linear accelerators

In a linear accelerator, one can use standing-wave cavities as described above or a travelling electromagnetic wave to accelerate electrons. Of course, the travelling wave needs to be propagating in a waveguide-like structure in order for the electric field to have a component along the direction of travel. Then one can inject electrons to sit on the crest of that travelling wave and gain energy as indicated in the cartoon in Fig. 3.7. A schematic sketch of an accelerating structure is shown in Fig. 3.8, where $2a$ is the diameter containing the beam and $2b$ is the outer diameter of the waveguide.

¹¹The breakdown mechanisms are different for conducting and superconducting cavities. For superconducting cavities, the hard limit is set by the fact that if the magnetic field at the surface becomes too large, the superconductor will return to the normal resistive state (‘quench’). In practice, no useful superconducting cavities have been made with gradients above 50 MV m⁻¹.

¹²There is currently some discussion about ideas for very large circular colliders, including the e^+e^- option. If such a machine is ever built, it will certainly be very expensive!

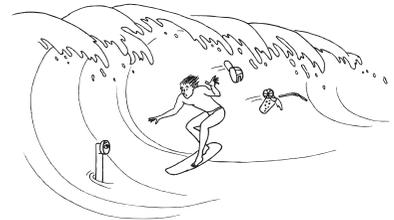


Fig. 3.7 The principle behind travelling-wave acceleration.

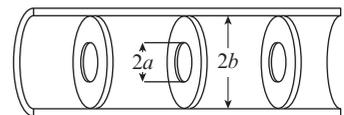


Fig. 3.8 Disk-loaded accelerating structure.

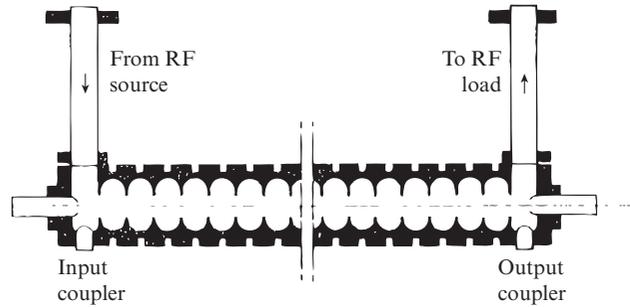


Fig. 3.9 Travelling-wave accelerating structure.

There are two approaches to particle acceleration. One is based on the use of cavities with short accelerating gaps (see eqn 3.14). An alternative approach uses a waveguide structure in which we have a travelling wave. However, in a smooth waveguide, the phase velocity is always larger than c and therefore cannot be used for particle acceleration. One approach to this problem is to insert discs inside the structure that are used to adjust the phase velocity of the travelling wave. The radii a and b and the distances between discs are chosen such that the phase velocity of the wave equals the electron velocity. They need to be changing along the structure as electrons are being accelerated. But once the electron speed becomes very close to the speed of light, there is no need to change the geometry of the structure. A more realistic sketch of an accelerating structure is shown in Fig. 3.9. An RF wave produced by a klystron enters and leaves each cavity to be absorbed outside the cavity. If instead of being absorbed the wave is reflected at the end of the cavity, a standing wave will be created that could also be used to accelerate electrons.

3.1.6 Klystrons

This section gives a very brief and simplified idea of how klystrons work. A DC high voltage is first used to accelerate a continuous electron beam. The electron beam enters an RF cavity to which RF power is delivered at a resonant frequency. This causes the velocity of the electron beam to become modulated. The electrons enter a drift region in which the velocity modulation is translated into spatial modulation (bunching). Finally, the electron bunches enter another RF cavity called the ‘catcher region’. They enter out of phase with the RF, so they are decelerated and their kinetic energy is converted into RF energy. The RF wave is then guided by a waveguide to the accelerating RF structure.

3.2 Beam optics

3.2.1 Magnetic lenses

To guide and focus beam particles along the reference trajectory (which may not be a straight line), one needs magnets. In a synchrotron, for example, which has a circular geometry, one needs dipole magnets

providing a vertical magnetic field (the accelerators are constructed in the horizontal plane) to bend the trajectories of beam particles so they stay inside the beam pipe (with a good vacuum inside), close to the reference trajectory. But the vertical magnetic field is not enough. For example, it does not constrain particle movement along the field direction, so any vertical component of the momentum, however small, would result in beam particles eventually escaping from the accelerator. One needs to have an arrangement of magnetic fields such that beam particles are effectively confined as if they were in a potential well that prevents them from going too far away from the reference trajectory. Quadrupole magnets are needed for this, together with other magnets for fine tuning. Only dipoles and quadrupoles will be considered here.

A schematic view of a dipole magnet is shown in Fig. 3.10. The beam pipe, a continuous vacuum chamber, runs through the yoke gap, where a magnetic field B is created by electric currents in the two coils. A ‘warm’ iron yoke can be used for fields up to about 2 T. To avoid iron saturation effects and achieve higher fields, one needs superconducting dipoles (see Section 3.3). The dipole bending strength for a particle with momentum p and charge q is then given by the inverse of the radius of curvature ρ :

$$\frac{1}{\rho} = \frac{qB}{p} \simeq 0.3 \frac{B [\text{T}]}{p [\text{GeV}/c]} \quad \text{for } q = e = \text{the electron charge} \quad (3.16)$$

In terms of the gap height h , the number of windings $n/2$ and the current I in each coil,

$$B = \frac{\mu_0 n I}{h} \quad (3.17)$$

where μ_0 is the permeability of free space.

A schematic cross section of a quadrupole magnet is shown in Fig. 3.11. Four pairs of coils, with n windings and current I in each coil,¹³ create a magnetic field with components $B_x = -gz$ and $B_z = -gx$ in the horizontal and vertical directions, respectively, where

$$g = \frac{2\mu_0 n I}{R^2} \quad (3.18)$$

and R is the distance shown in Fig. 3.11. The corresponding components of the Lorentz force acting on a particle with speed v are

$$F_x = qvB_z = -qvgx \quad \text{and} \quad F_z = -qvB_x = qvgz \quad (3.19)$$

The important point to note is that in the vertical plane (containing the origin, i.e. the reference trajectory point) the force is acting away from the origin, while in the corresponding horizontal plane it acts towards the origin. One says that the quadrupole is focusing in one plane and defocusing in another plane, perpendicular to the first. There is a complete

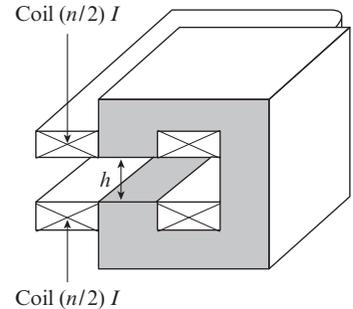


Fig. 3.10 Schematic view of a dipole magnet.

¹³This equation assumes that we are not using a warm iron core magnet, so it is only valid for superconducting quadrupoles.

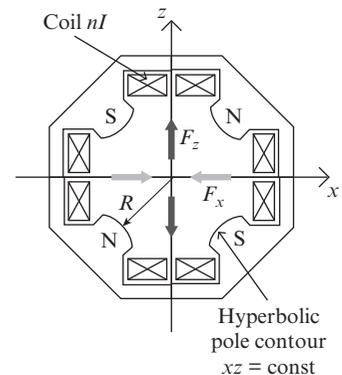


Fig. 3.11 Schematic view of a quadrupole magnet.

analogy with geometrical optics, and for a quadrupole of length l , with quadrupole strength k , one can define its focal length f :

$$\frac{1}{f} = k, \quad \text{where } k [\text{m}^{-2}] = \frac{qg}{p} \simeq 0.3 \frac{g [\text{T/m}]}{p [\text{GeV}/c]} \quad \text{for } q = e \quad (3.20)$$

For the HERA proton ring, $k \simeq 0.033 \text{ m}^{-2}$, $l \simeq 1.9 \text{ m}$, and $f \simeq 16 \text{ m}$. If $f \gg l$, the quadrupole can be treated as a thin lens irrespective of the absolute value of l .

A thin lens of focal length f_1 and another thin lens of focal length f_2 arranged as a doublet of lenses separated by a drift tube of distance l form a focusing doublet with an effective focal length f (see Exercise 3.4b) given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{l}{f_1 f_2} \quad (3.21)$$

If one lens is focusing in the horizontal plane (f positive) and one defocusing (f negative), then we can arrange for the effective focal length of the system to be positive. Such a quadrupole doublet, with a drift space between the two quadrupoles, can be arranged to give focusing in both the horizontal and vertical directions. For example, let $f_1 = f_Q$ and $f_2 = -f_Q$; then the effective focal length is given by $f = f_Q^2/l$ in both the horizontal and vertical dimensions. This is the idea behind so-called ‘strong focusing’, in which focusing and defocusing quadrupoles are arranged in doublets with a dipole inside each doublet.¹⁴ A structure like this is called a FODO cell, as sketched in Fig. 3.12. A FODO cell focuses beam particles in both planes. FODO cells are put together one after another as a periodic structure along the whole ring of a synchrotron. Calculations of particle trajectories inside such a structure can be performed using the same techniques as used in geometrical optics; hence this aspect of accelerator physics is called ‘beam optics’. The concept of using repeating structures of FODO cells is called strong focusing and it keeps the transverse dimensions of the beam relatively small all the way around the ring, allowing the use of relatively small beam pipes and magnets. Before strong focusing was discovered, synchrotrons used ‘weak focusing’, which resulted in much larger beam pipes. Strong focusing was another key development that allowed the construction of very high-energy synchrotrons at an affordable price.

¹⁴From the focusing perspective, the dipole magnet acts as a drift tube. Note that if l is too short, then the effective focal length will become too long.

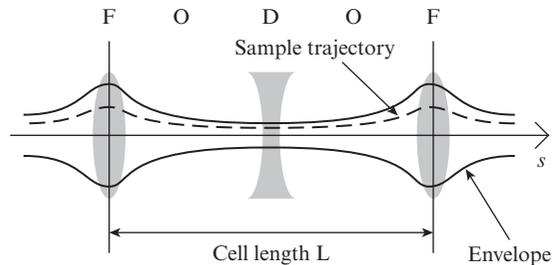


Fig. 3.12 FODO cell.

3.2.2 Beam trajectories and phase space

The motion of beam particles is described in a curvilinear coordinate system, as sketched in Fig. 3.13. In a first, linear, approximation, particle motion in each of the three space directions (longitudinal s , vertical z , and horizontal x) can be considered separately. The transverse phase space is split into two 2-dimensional phase spaces. Considering, for example, the vertical direction, we have the z -coordinate and p_z -component phase space. As indicated in Fig. 3.14, the velocity z component can be described as a product of the angle with respect to the s direction and the speed along s , which is approximately the speed of light. So, effectively, for a given and constant Lorentz γ factor, what matters is the angle, and the (z, p_z) phase space can be replaced by $(z, z' = dz/ds)$ phase space. A similar argument applies in the other transverse direction.

The strength of a quadrupole is defined by

$$k = \frac{1}{B\rho} \frac{dB_z}{dx} \quad (3.22)$$

where ρ is the radius of curvature of the reference trajectory.¹⁵ The equations of motion are then Hill's equations

$$z'' + k(s)z = 0 \quad (3.23)$$

$$x'' - \left[k(s) - \frac{1}{\rho^2} \right] x = \frac{1}{\rho} \frac{\Delta p}{p} \quad (3.24)$$

We will only consider solutions for z and z' (or for the horizontal phase space for $\Delta p = 0$, i.e. for the nominal momentum and at the limit of $\rho \rightarrow \infty$). This differential equation is reminiscent of simple harmonic motion, but $k(s)$ is not a constant; it is a periodic function that defines the focusing strength at any point along the ring (eqn 3.22). It is obvious that if there were no focusing and $k(s) = 0$, then beam particles could escape unimpeded. If $k(s)$ were constant around the ring, then the solution would be simple harmonic motion. This suggests the use of oscillatory trial functions similar to those for simple harmonic motion:¹⁶

$$\begin{pmatrix} z \\ z' \end{pmatrix} = \begin{pmatrix} \sqrt{\epsilon} \sqrt{\beta(s)} \cos[\varphi(s) - \varphi_0] \\ -\frac{\sqrt{\epsilon}}{\sqrt{\beta(s)}} \{ \sin[\varphi(s) - \varphi_0] + \alpha(s) \cos[\varphi(s) - \varphi_0] \} \end{pmatrix} \quad (3.25)$$

The initial conditions determine the values of ϵ and φ_0 . The function $\beta(s)$ defines the amplitude modulation, which varies because of the changing focusing strength around the ring. From our trial solution, we can also derive the relationship between the function $\beta(s)$ and the magnetic focusing $k(s)$. It is convenient to define $\beta = \omega^2$, and we then find¹⁷

$$\omega'' - \frac{1}{\omega^3} + \omega k(s) = 0 \quad (3.26)$$

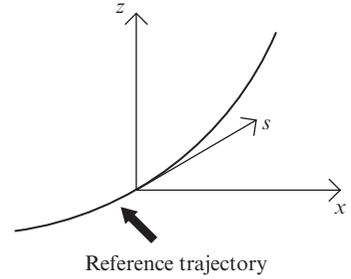


Fig. 3.13 Curvilinear coordinate system along the reference trajectory.

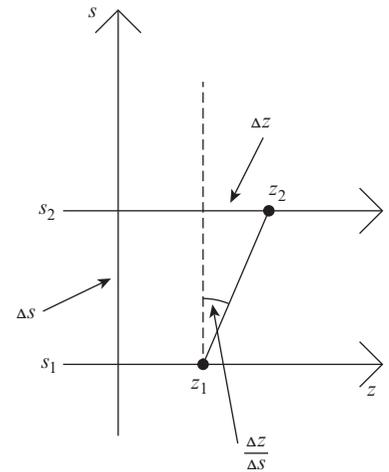


Fig. 3.14 A particle going from z_1 to z_2 in the vertical direction.

¹⁵In general, as we go around a ring (s), the magnetic focusing will vary, so we write $k(s)$ to remind ourselves that k is not a constant.

¹⁶See Exercise 3.3 for a justification.

¹⁷See Exercise 3.3.

¹⁸See Wille in Further Reading.

¹⁹See Exercise 3.3.

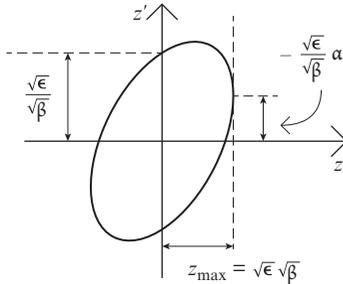


Fig. 3.15 The phase-space ellipse in the (z, z') plane.

²⁰The horizontal and vertical Q values are not related to the Q-values of the RF cavities discussed in section in 3.1.3.

²¹We will see how to evade Liouville's theorem when we consider stochastic cooling in Section 3.6.

In principle, this allows us to determine $\omega(s)$ and hence $\beta(s)$ if we know $k(s)$. However, this is not practical and matrix methods are used to determine $\beta(s)$.¹⁸ The phase advance function $\phi(s)$ is also determined by the focusing. The resulting oscillations about the reference trajectory are called betatron oscillations. The amplitudes of z and z' are written using the amplitude function $\beta(s)$ and the emittance ϵ of the trajectory (at the moment, we are talking only about one particle). The optical function $\alpha(s) \equiv -\beta'(s)/2$ and the phase function $\varphi(s)$ are related to the $\beta(s)$ function by¹⁹ $\varphi'(s) = 1/\beta(s)$.

At a given point s along the reference trajectory, a particle has coordinates (z, z') in the vertical phase space. After one revolution, it will come back to the same s but with different coordinates (z, z') . After many revolutions the particle's coordinates (z, z') will trace an ellipse in the (z, z') phase space, as shown in Fig. 3.15. Integrating the phase function around the circumference C of the accelerator gives

$$\int_s^{s+C} d\varphi = 2\pi Q_z \quad (3.27)$$

where Q_z is known as the betatron tune, the number of betatron oscillations for a particle going around the accelerator once (in this case the vertical tune Q_z ; similarly, there is a horizontal tune Q_x).²⁰ If there are any small imperfections in the ring, we need to avoid particles crossing these imperfections at the same betatron phase in each revolution—otherwise the beam would rapidly ‘blow up’. Therefore, integer values of the betatron tune should be avoided. More generally, tunes satisfying

$$\nu Q_x + \mu Q_z = \xi \quad (3.28)$$

with ν , μ , and ξ integers must be avoided.

The area of the (z, z') ellipse is $\pi\epsilon$, so, up to the factor of π , the emittance is the volume of the two-dimensional phase space (to be more precise, here ϵ_z and similarly ϵ_x for the horizontal phase space). Liouville's theorem states that under the action of conservative forces, the volume of beam phase space is conserved.²¹ Therefore, on moving from one point s to another along the reference trajectory, one would get another ellipse but with the same area. So the transverse motion of a particle can be visualized as an ellipse of fixed area changing its shape depending on the location in the accelerator. If there is another particle in the accelerator that is described by the same equations of motion but has a different initial phase φ_0 , then the motion of that particle will be given by the same ellipse, since a different value of φ_0 simply corresponds to another point on the same ellipse. So, in fact, one ellipse describes a family of trajectories, not just a single trajectory. From the algebraic point of view, this family of trajectories is described by the amplitude function $\beta(s)$ and the emittance ϵ (see eqn 3.25). But the emittance is constant for ‘coasting’ (no acceleration) beam particles, which means that we have reduced the problem from two dimensions (z, z') to one dimension (β) . On inspecting Fig. 3.15, we can see that

the amplitude function $\beta(s)$ is the ratio of the beam width to the on-axis angular spread.

Each ellipse represents a family of particles, and the whole ensemble of beam particles consists of many of these families and ellipses. How do we represent the whole ensemble? Considering the vertical phase space (the same argument applies for the horizontal one), a particle beam injected into an accelerator is characterized by initial conditions equivalent to a cluster of points in the (z, z') phase space, centred about the reference trajectory $(0, 0)$. We choose an ellipse that closely surrounds this cluster, and this represents the ‘edge’ of the beam. By convention, the ellipse should contain 95% of particles. Then we follow this ellipse through the accelerator; the ellipse, the corresponding amplitude function $\beta(s)$, and the beam emittance ϵ represent the properties of the whole beam.

Using Liouville’s theorem, we see that as long as the beam is not accelerated, the (z, z') and (z, p_z) phase spaces are equivalent. But once the beam is accelerated, Liouville’s theorem applies only to the proper phase space (z, p_z) , and only the normalized emittance $\epsilon_N = \beta\gamma\epsilon$ (here β is the speed and γ is the Lorentz factor) is conserved. The volume of the (z, z') phase space shrinks with the momentum p as $1/p$ and consequently the beam width and the beam angular divergence shrink during acceleration, each as $1/\sqrt{p}$; so a higher-energy beam fits into a smaller-diameter beam pipe. This explains why high-energy accelerators require chains of lower-energy accelerators; each accelerator in the chain reduces the emittance sufficiently to allow the beam to have sufficiently small emittance to fit into the next accelerator in the chain. In principle, this accelerator chain could be eliminated if the beam pipe of the high-energy accelerator were sufficiently large; however, this would increase the size and hence the cost of the magnets.

An example of such a chain is at the LHC [73], where the source of protons is a bottle of hydrogen gas. A high voltage is used to strip electrons to provide the protons. A linear accelerator (Linac 2) accelerates the protons to an energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which accelerates the beam to 25 GeV. Protons are then injected into the Super Proton Synchrotron (SPS), where they are accelerated to 450 GeV before injection into the LHC.

3.3 LHC dipole magnets

The LHC superconducting dipoles use conductors made from a niobium–titanium (NbTi) alloy.²² For these dipoles [73], which are capable of generating a magnetic field $B = 8.3$ T, the current required is $I = 11.85$ kA. This requires the NbTi superconductor to be cooled to a temperature of 1.9 K using superfluid helium.²³ In a type I superconductor, the current flows only on the surface, not in the bulk, which limits the useful magnetic field. Therefore, high-field superconducting magnets rely on

²²NbTi is the only low-temperature superconductor that is ductile, and hence most existing superconducting magnets are based on this alloy, although there is interest in the niobium–tin alloy Nb₃Sn, which might be able to produce larger magnetic fields. This would allow the option of an upgrade to the LHC to reach a CMS energy of 33 TeV. Nb₃Sn superconductors are used in some high-field MRI magnets.

²³This has the advantage of benefiting from the remarkable thermal properties of superfluid helium, which has an effective thermal conductivity orders of magnitude better than copper.

type II superconductors, in which magnetic fluxoids can penetrate the volume. When there is a changing magnetic field in a superconductor, this will cause screening currents to flow. These are similar to eddy currents, but as there is no resistance they do not decay with time. This magnetization appears as an unwanted error in the field produced by the magnet. The magnetization is proportional to the diameter of the wire carrying the current. When the magnetic field is changing with time, as happens when the beam energy is being ramped up from injection energy, an additional magnetization is created from the flow of current between neighbouring filaments. Therefore, a useful superconducting cable has to be made from a very large number of very small filaments wound as ‘twisted pairs’ to minimize the magnetization. The cable for the LHC dipole magnets [73] is based on $6\ \mu\text{m}$ -diameter filaments. The filaments are embedded in a copper matrix for mechanical support.²⁴ Each strand is made from 6300 filaments and is 0.825 mm in diameter, and 36 strands are then used to make a cable, as shown in Fig. 3.16.

²⁴Copper is effectively an insulator when the NbTi is in the superconducting state.

²⁵This type of cable is called Rutherford cable because it was developed at the UK Rutherford Laboratory, and it is used in all high-field superconducting magnets. The cable is used in MRI scanners, so this is probably one of the most important but least known spin-offs from particle physics research.

For the same reasons as discussed above, it is important to minimize the flux linkage between wires. Twisting wires around each other as in a conventional cable is not sufficient, because the inner (outer) wires remain inside (outside). The wires need to be fully transposed; i.e. every wire must change places with every other wire along the length of the cable so that, averaged over the length, no flux is enclosed.²⁵

To create a perfect dipole field, a distribution of current density varying as $\cos\phi$ around the beam pipe would be required. However, a very nearly uniform dipole field near the centre of the beam pipe is created by blocks of superconductor arranged in the geometry shown

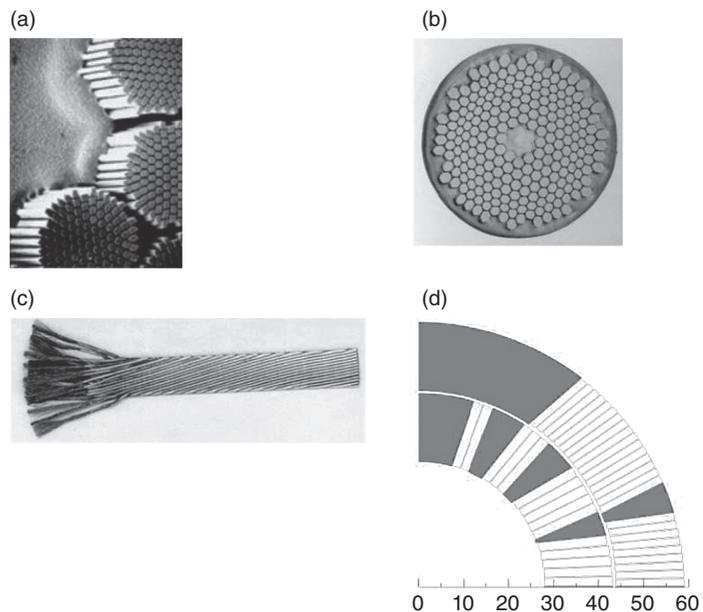


Fig. 3.16 Filaments (a), strands (b), and cable (c) of the type used for the LHC superconducting magnets, and a cross-section of one-quarter of the coils used in a main dipole (d). Note that the superconducting cable to create the dipole field is placed in the small outlined boxes in (d). From [73].

in Fig. 3.16(d). The currents in the blocks are optimized to produce a uniform magnetic field.

3.3.1 Engineering design details

As there is insufficient space in the LHC tunnel for separate magnets for each beam, a ‘two-in-one’ magnet was designed in which the two magnetic volumes are inside a common cryostat [73], as illustrated in Fig. 3.17.²⁶ This magnet design is an amazing engineering tour de force, as the figure shows. In comparison, the dipole magnets for a $p\bar{p}$ collider are straightforward.

²⁶This also allowed for significant cost savings compared with having two separate magnets and cryostats.

The two beam pipes containing the counter-circulating proton beams are in the centre, surrounded by their respective dipole bending magnets with fields orientated to bend the two separate positively charged particle beams in opposite directions. The superconducting cable is held in place by non-magnetic ‘collars’ of austenitic steel, which can withstand a magnetic force of about 400 tonnes per metre of dipole. At nominal operation, the energy stored in each of the 1232 LHC dipoles is 6.93 MJ. If one small region of the superconductor becomes

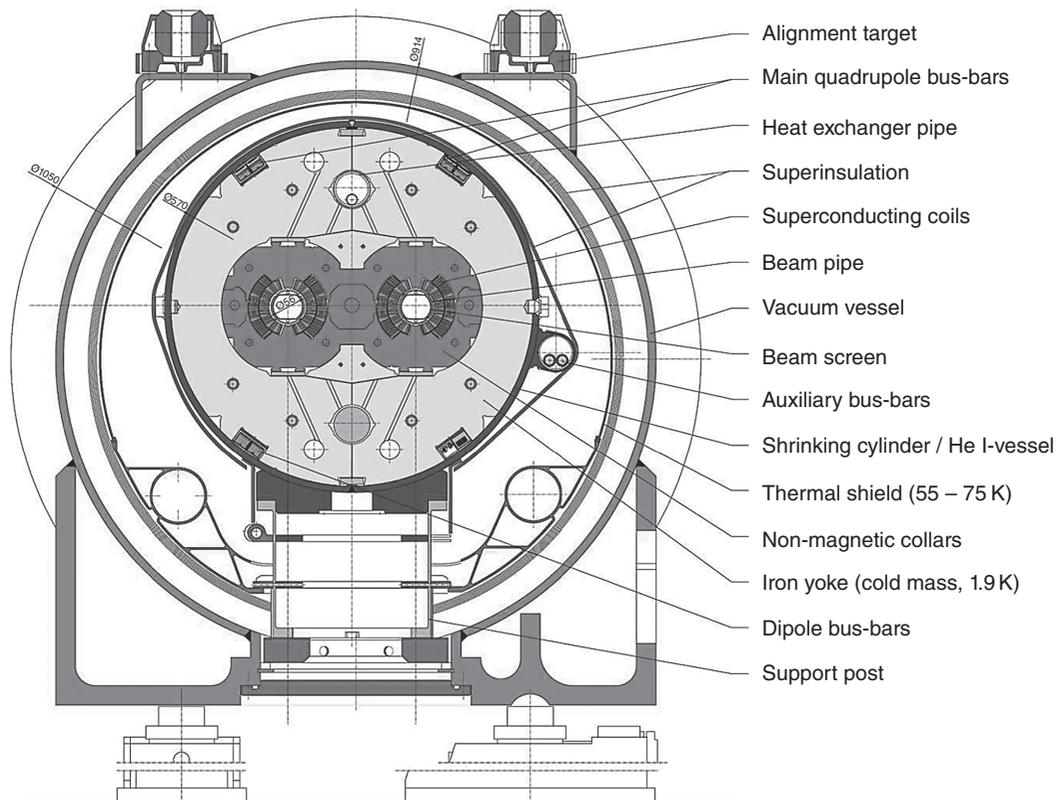


Fig. 3.17 Cross-section of an LHC dipole in its cryostat. From [73].

non-superconducting (called a quench) for any reason, there will be Joule heating, which will increase the resistance, and hence there is the possibility of a catastrophic runaway, which would destroy the magnet. Therefore, sophisticated quench detection and protection systems are essential. Quenches can be detected by the extra ‘*IR*’ voltage drop. Once a quench is detected, a ‘quench heater’ is operated to force the entire magnet to become non-superconducting and the energy is transferred to a large ‘dump’ resistor.²⁷

²⁷An additional source of danger is the electrical connection between the dipoles. This has a tiny but non-zero resistance. If this resistance is too large, this can also lead to catastrophic thermal runaway, as occurred on 19 September 2008 and led to extensive damage. Many improvements have been made since then to prevent this type of problem.

The energy stored in the two beams at nominal operation is 362 MJ, so they have sufficient energy to destroy large parts of the LHC machine and the detectors. Therefore, many beam loss monitors are installed in the machine, and if the rates rise above a threshold, kicker magnets are operated to deflect the beams out of the ring towards a beam dump [73]. The beam dump must be able to dilute the peak energy density of the beam before it is absorbed. At the LHC, this is done by ‘spraying’ the beam in a spiral pattern as it enters the dump.²⁸

²⁸Somewhat analogous to what one does with a garden hose to avoid damaging a sensitive plant.

3.4 Colliders and fixed-target accelerators

The centre-of-mass energy in a symmetric collider, with each beam having energy E , is simply given by $\sqrt{s} = 2E$. For a fixed target collision, with a beam energy E and a target mass m (assuming $E \gg m$), we can show²⁹ that

²⁹See Exercise 3.1.

$$\sqrt{s} = \sqrt{2mE} \quad (3.29)$$

Therefore, colliders have an obvious advantage in maximizing the centre-of-mass energies over fixed-target experiments. Although this was realized a long time ago, the challenge of achieving a useful interaction rate was formidable. The interaction rate for a given physics process depends on the luminosity (see Section 3.5). In a fixed-target geometry, it is only necessary for one high-intensity beam to collide with a block of matter to achieve very high luminosities. For a collider, this is far more challenging because we need two intense beams, which both have to be focused to very small transverse dimensions at the interaction points to achieve a useful luminosity. In a fixed-target accelerator, the beam must be kept for a few seconds before it is extracted. However, in a collider, it takes time to ‘fill’ the machine with sufficient numbers of particles before they are accelerated to the peak energy, and then the beams have to be kept for several hours while data are taken. This obviously puts much more significant demands on the quality of the accelerator. We must avoid the dangerous resonances and ensure that imperfections, which can increase the emittances of the beams, are kept to a minimum. We also need an extremely high vacuum to minimize beam losses and backgrounds in the detector.³⁰ The defocusing effects of one beam on the other must also be kept under control.

³⁰At the LHC, the pressure has to be kept below about 10^{-7} Pa.

Some of these issues are common to all types of colliders. We always want higher luminosity, and how to achieve this is discussed in Section 3.5. The special issue for circular e^+e^- colliders is synchrotron radiation (see eqn 3.15), which puts a practical limit on the beam energy. Therefore, large e^+e^- colliders like LEP have simple (and cheap) magnets but require very efficient RF cavities—which means superconducting cavities (see Section 3.1). Hadron colliders do not suffer from significant synchrotron radiation losses, so we can have much higher beam energies. To optimize the beam energy for a given cost, we need to use the highest magnetic field possible for the dipole bending magnets. The critical technological challenge is the industrial-scale production of very high-quality superconducting magnets (see Section 3.3).

In an e^+e^- collider, the energy defines which processes can be studied³¹ and which particles can be created or discovered. In a hadron collider such as the LHC, there is a more subtle interplay between energy and luminosity. We are really interested in the rates for processes at the parton level, and the partons only carry a fraction of the momentum of the protons, so the energy reach of a hadron collider depends crucially on both energy and luminosity. Therefore, there is some complementarity between the very clean physics that can be performed at e^+e^- colliders, compared with the higher energy reach of hadron colliders, in which the event reconstruction is more complicated because, in addition to the interesting parton–parton collisions, there are always interactions of the remaining ‘spectator quarks’.

The HERA collider was a special case because it used $e^\pm p$ collisions. This required a high-energy proton ring (HERA I 820 GeV, HERA II 920 GeV) and a much lower-energy electron ring (27.5 GeV) to minimize synchrotron radiation.³² As will be discussed in Chapter 9, this gave the most precise determination of the quark and gluon distribution functions. These are required for all calculations of cross sections at hadron colliders like the LHC; the details will be covered in the same chapter.

The LHC uses pp collisions, but earlier hadron colliders used $\bar{p}p$ collisions. The important advantage of $\bar{p}p$ colliders is that the two beams can be contained in the same beam pipe. The critical issue for these colliders was how to produce intense and low-emittance \bar{p} beams. This will be discussed in Section 3.6. Colliders have also been operated using heavy ions, but these will not be discussed in this book.

3.5 Luminosity

The two most important numbers in experimental (accelerator) particle physics are the energy and the luminosity of an accelerator. The luminosity \mathcal{L} translates a cross section σ into the rate R of produced events:³³

$$\frac{dR}{d\Omega} = \mathcal{L} \frac{d\sigma}{d\Omega} \quad (3.30)$$

³¹Energy conservation implies that for a beam energy E , in a symmetric collider, the most massive particle that can be pair-produced will have mass $m = E$. In general, the sum of the masses of the final-state particles must be less than the centre-of-mass energy ($2E$).

³²The ep CMS energies were 300 and 318 GeV, respectively.

³³Or observed events, if detector effects are included.

where Ω is the solid angle for the angular differential cross section. In fixed-target experiments,

$$\mathcal{L} = n\rho l \quad (3.31)$$

where n is the number of particles per second in the beam (typically 10^{12} s^{-1}), ρ is the density of target particles, and l is the target length (typically, $\rho l \simeq 10^{23} \text{ cm}^{-2}$), giving a luminosity $\mathcal{L} \simeq 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, which is large in comparison with that at a high-energy collider.³⁴

We begin by deriving useful formulae for the luminosity and then discuss how to optimize it.

We start with the simple scenario sketched in Fig. 3.18, with two rectangular bunches of particles colliding head on. They contain respectively n_1 and n_2 randomly distributed particles. A is the transverse area of the wider bunch and there are b bunches in each beam, with frequency of revolution f . Then

$$\mathcal{L} = \frac{n_1 n_2}{A} b f \quad (3.32)$$

In a collider, n_2/A corresponds to the fixed-target ρl ; it is the number of target particles per unit area. If particles are distributed in colliding bunches not randomly but according to normalized density distributions ρ_1 and ρ_2 , then

$$\mathcal{L} = b f n_1 n_2 \int_S \rho_1 \rho_2 dS \quad (3.33)$$

where S is the transverse space. Introducing beam currents $I_1 = n_1 e f b$ and $I_2 = n_2 e f b$, where e is the electron charge,

$$\mathcal{L} = \frac{I_1 I_2}{e^2 b f} \int_S \rho_1 \rho_2 dS \quad (3.34)$$

Assuming Gaussian distributions with $\sigma_x = \sigma_z = \sigma$ for simplicity,³⁵

$$\rho_{1,2} = \frac{1}{2\pi\sigma_{1,2}^2} \exp\left(-\frac{r^2}{2\sigma_{1,2}^2}\right) \quad (3.35)$$

we get

$$\mathcal{L} = \frac{I_1 I_2}{e^2 b f} \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} \quad (3.36)$$

where $2\pi(\sigma_1^2 + \sigma_2^2)$ represents an effective area.

In order to illustrate how this relates to the accelerator parameters, we will assume that the vertical and the horizontal emittances are equal, $\epsilon_x = \epsilon_z = \epsilon/\pi$, and that the horizontal and the vertical beta functions are equal, $\beta_x = \beta_z = \beta^*$. We also assume that the two beams have equal emittance. The asterisk in β^* is the common symbol to indicate that the function is calculated at the interaction point (IP). As

³⁴This simple formula assumes that the target is ‘thin’, i.e. that the probability that an individual beam particle interacts in the target is much less than one.

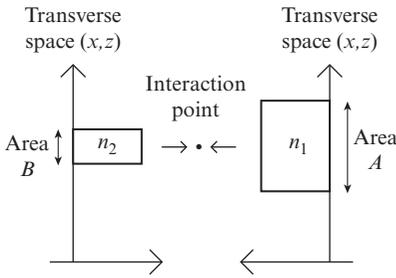


Fig. 3.18 Simple geometry for derivation of the luminosity formula. Symbols are explained in the text.

³⁵This is not very realistic, since typically the horizontal beam size is bigger than the vertical one.

$$\sigma^2 = \epsilon\beta^* \quad (3.37)$$

$$\mathcal{L} = \frac{I_1 I_2}{4\pi e^2 b f \beta^* \epsilon} \quad (3.38)$$

Equation 3.38 gives the best guide for understanding how to maximize the luminosity for a collider. First, it is clear that increasing the beam currents I_1 and I_2 is desirable. If only a limited number of protons can fit in one bunch, then it is advantageous to increase the number of bunches. However, the beam currents cannot be increased indefinitely, because each beam exerts electromagnetic forces on the other beam at each IP. The net effect of one bunch on a counter-rotating bunch is similar to that of an additional quadrupole magnet, and therefore changes the horizontal and vertical Q values.³⁶ This is very dangerous because, even if the operating point of the machine is away from integer resonances (see eqn 3.28), such a tune shift can push the beams too close to a resonance, resulting in very rapid beam loss. This beam–beam tune shift thus limits the ultimate luminosity that can be achieved in a hadron collider.³⁷

Note the rather counter-intuitive result of eqn 3.38 that if the beam currents are at the beam–beam limit, then the luminosity can be increased by decreasing the number of bunches. However, the optimization of the number of bunches must also respect practical constraints imposed by the detectors. If the number of bunches is reduced, the number of collisions per bunch crossing will increase. Therefore, collisions with one interesting event will also contain a background of many other ‘minimum-bias’ collisions, which are effectively a noise source.³⁸ There is therefore a trade-off between maximizing luminosity and having clean enough events to be useful—and there is no perfect solution. At the LHC design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, there will be approximately 25 collisions per bunch crossing (every 25 ns).

The next parameters to optimize are the emittances. These depend on the quality of the proton source. Although Liouville’s theorem predicts the conservation of beam phase space, any imperfections can increase the emittances. Finally, one can increase the luminosity by decreasing the values of β^* . This is achieved by using very strong focusing quadrupole magnets, which will usually be superconducting to achieve the highest field gradients. The consequence of reducing the transverse beam size at the IP is that the beam divergences will increase. This limits how far β^* can be reduced before the beam losses from particles hitting the beam pipe become unacceptable. This, in turn, implies that there is an advantage in bringing the quadrupole magnets closer to the IP, but as this will reduce the space for detectors in the forward region, the trade-off will depend on the particular physics being studied.

³⁶This effect is called the ‘beam–beam’ tune shift. See Wille in Further Reading for a full explanation. The horizontal and vertical Q values are not related to the Q -values of the RF cavities discussed in section in 3.1.3.

³⁷The situation is different in e^+e^- colliders because of the beam ‘cooling’ from synchrotron radiation.

³⁸How to cope with this is covered in more detail in Section 13.2.

3.6 $\bar{p}p$ colliders

Hadron colliders can use two separate p beams as in the LHC. However, this requires two separate magnetic fields for the counter-rotating

³⁹They also demonstrated to sceptical physicists that very clean results could be obtained in this difficult environment.

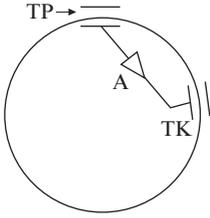


Fig. 3.19 Schematic view of a synchrotron with transverse pickup (TP), fast amplifier (A), and transverse kicker (TK) electrodes for stochastic cooling.

⁴⁰Hence the name ‘stochastic cooling’.

p beams. With counter-rotating beams of particles and antiparticles, the electric and magnetic fields are automatically correct for both beams, so only one ring is required. This enabled the relatively cheap conversion of proton synchrotrons at CERN and Fermilab into $p\bar{p}$ colliders. The use of $p\bar{p}$ colliders was very important as it led to the discovery of the W and Z bosons and the top quark, as well as providing the most precise measurement of the W mass.³⁹

The big challenge for $p\bar{p}$ colliders was to produce very intense, low-emittance, beams of antiprotons, which was achieved using the technique of stochastic cooling [132]. Liouville’s theorem prevents a reduction in beam phase space, but it is based on the assumption that the beam is continuous, whereas a real beam is composed of a finite number of discrete particles. Consider first the extreme case of a single particle in a beam; we can detect its transverse position using a beam pickup electrode at one place in the ring, which generates a signal proportional to the displacement about the central orbit. The signal is fed across the ring via an amplifier to two deflecting plates as shown in Fig. 3.19.

The betatron function at the deflecting plates is an odd multiple of $\pi/2$ out of phase with that of the pickup, so that a deviation in position from the reference orbit is converted to a difference in angle. The shorter path for the cable compared with the particle trajectory compensates for the difference in speeds of the particles and the electrical signals as well as the delay in the amplifier. Therefore, a suitable voltage pulse can be used to make a correction to bring the particle back to the central orbit.

In a real beam with a large number of particles, this cooling technique works on a statistical basis.⁴⁰ If the speed of the amplifier were sufficiently fast, then each individual particle in the beam would have the correction applied to bring it back to the central orbit. Therefore, the cooling works better the shorter the sample time of the amplifier, since this determines how many particles are affected. Hence, the cooling rate depends on the bandwidth of the amplifier, W . As stochastic cooling requires the detection of fluctuations, it works faster for smaller numbers of beam particles, N , and the cooling time $\tau \approx N/2W$. For useful bunches, we need $N \approx 10^{12}$, and with achievable bandwidth amplifiers this leads to a cooling time of the order of 1 day.

It is also necessary to provide momentum cooling to reduce the spread in momenta. A pickup electrode can be used to measure the revolution frequencies of particles, and the signal is fed into an amplifier via a filter. The filter eliminates any signal for particles with the correct frequency (i.e. momentum), and higher frequencies give a phase shift of π . This filtered signal can then be fed into an accelerating cavity. Again this system would work perfectly for individual particles and works on a statistical basis for beams with a finite number of particles.

3.6.1 CERN $p\bar{p}$ collider

The antiprotons were produced by collisions of 26 GeV/ c protons from the CERN PS with a copper target. Some of the produced antiprotons

with momenta $\sim 3.5 \text{ GeV}/c$ were collected by a large-aperture low-energy ring called the antiproton accumulator. The pulse of antiprotons was first cooled and then moved to the side of the aperture, where an intense stack of antiprotons was built up, allowing a new injection of antiprotons every 2.2 s. The antiprotons were accumulated and cooled using stochastic cooling for about one day and then injected back into the PS and accelerated to $26 \text{ GeV}/c$; they were then injected into the SPS together with counter-rotating bunches of protons. The protons and antiprotons were then accelerated to an energy of 315 GeV (initially 270 GeV), and a run would last until sufficient antiprotons had been accumulated or the beams were lost. The peak luminosity achieved was $\sim 2 \times 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$.

3.6.2 Tevatron $\bar{p}p$ collider

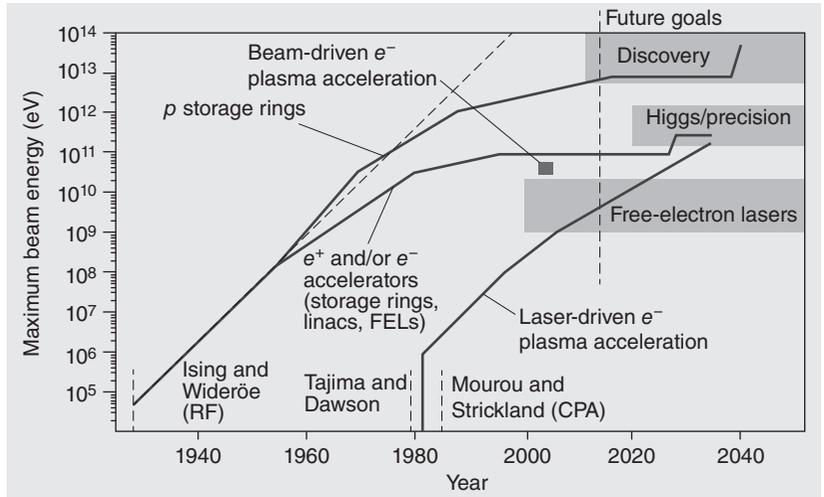
Similar principles were applied to the Tevatron. For the second phase of the Tevatron (Run 2), a 150 GeV synchrotron called the main injector was built to provide higher yields of antiprotons. The acceptance of antiprotons and the cooling were split between two separate machines. After cooling, the protons and antiprotons were re-accelerated in the main injector and then injected into the superconducting Tevatron and accelerated to an energy of 0.98 TeV . The peak luminosity achieved was $\sim 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$.

3.7 Future accelerators

The RF technology for accelerating particles has been developed over many years, providing accelerators with steadily growing energies and luminosities. Proposed new accelerators, such as the International Linear Collider and the Future Circular Collider, are still based on RF technology. However, the size of these accelerators and their high cost demand a new, cheaper technology to allow the field of particle physics to move beyond current plans and also make applications affordable, for example as a source of ultrashort X-ray pulses to study the dynamics of chemical reactions. One such technology where progress in recent years has been particularly impressive is plasma acceleration. A plasma is an ionized gas such as hydrogen. By moving electrons away from quasistationary ions, it is possible to create electric fields up to three orders of magnitude larger than those obtainable using RF technology, thus allowing accelerators to be much smaller in size. A high-intensity laser pulse or a train of laser pulses or bunches of charged particles like electrons or protons can be used to separate electrons from ions, creating very high electric fields in plasmas. Electrons have been accelerated to energies above 4 GeV in a 9 cm -long accelerator in the BELLA laboratory at Berkeley,⁴¹ demonstrating the clear potential of this technology. The maximum beam energies reached or expected to be reached at future facilities are presented in Fig. 3.20, together with the longer-term prospects for plasma accelerators.

⁴¹Berkeley Lab Laser Accelerator Center.

Fig. 3.20 The Livingston plot of the maximum beam energy for conventional RF and plasma accelerators. Image credit: R. Assmann.



Chapter summary

- This chapter has given a brief introduction to the field of accelerator physics.
- We have seen how RF cavities are used to accelerate charged particles.
- Dipole magnets are required for circular machines, but quadrupoles are also essential for beam focusing. The oscillations of beam particles about their central orbit have been explained and the beam emittance defined.
- Superconducting cables are needed for the highest-energy synchrotrons such as LHC.
- Colliders are the best route to the highest-energy collisions, provided sufficient luminosity can be achieved. A brief explanation of luminosity and its optimization has been given.
- The way in which Liouville's theorem can be circumvented with the use of stochastic cooling has been reviewed. This was essential for the operation of $\bar{p}p$ colliders.

Further reading

- Wilson, E. (2001). *An Introduction to Particle Accelerators*. Oxford University Press. This is a very good introduction to the subject.
- Wille, K. (2000). *The Physics of Particle Accelerators: An Introduction*. Oxford University Press. This is a more advanced introduction to the subject.
- Bryant, P. and Johnsen, K. (1993). *The Principles of Circular Accelerators and Storage Rings*. Cambridge University Press.
- The Staff of the CERN Proton–Antiproton Project (1981). First proton–antiproton collisions in the CERN SPS collider. *Phys. Lett. B*, **107**, 306.

Exercises

(3.1) Use relativistic kinematics to derive eqn 3.29.

(3.2) (a) Using the definition in eqn 3.7, show by partial differentiation that

$$\eta_{\text{RF}} = \left(\frac{\partial f}{\partial v} \frac{\partial v}{\partial p} + \frac{\partial f}{\partial R} \frac{\partial R}{\partial p} \right) \frac{p}{f}$$

(b) Show that

$$\frac{\partial v}{\partial p} = \frac{1}{m\gamma^3}$$

(c) Combine the results from (a) and (b) to derive eqn 3.8.

(3.3) (a) Differentiate twice the trial solution of Hill's equation (3.25) in the z direction.

(b) Using the requirement that the coefficients of \sin and \cos must be identical, together with the results from (a), show that our trial solution to eqn 3.24 is valid provided the condition $\phi' = 1/\beta$ is satisfied.

(c) Verify eqn 3.26 by equating coefficients of the \cos terms in Hill's equation (3.24).

(3.4) (a) Write down the matrices in the vertical direction (z, z') for a thin focusing (F) and defocusing (D) lens and a drift tube (O) (i.e. with no magnetic forces).

(b) Use matrix multiplication to evaluate the matrix for the combination FOD and hence derive eqn 3.21.

(3.5) Consider the quadrupole magnetic field given by $B_x = -gz$ and $B_z = -gx$. Show that this field satisfies Maxwell's equations in free space. Sketch the resulting magnetic field lines. Hence sketch the magnetic forces acting on a positive particle travelling parallel to the beam axis (s). Consider the trajectory of such a particle after traversing such a thin quadrupole lens of length l . The particle has an initial horizontal coordinate $x = x_0$. Show that all such particles will have $x = 0$ after travelling a distance given by the focal length (eqn. 3.20). *Hint:* Assuming the lens is thin, you may neglect the change in x of the particle inside the quadrupole.

(3.6) A collider with two beams of unequal width has 3×10^{11} particles per bunch and two bunches in each beam, and the frequency of revolution is 1 MHz.

Particles are uniformly distributed in cylindrical bunches that move parallel to the cylinder axis. The radius of each bunch in the wider beam is 4×10^{-4} m. Calculate the luminosity.

(3.7) Describe a method to determine the luminosity in an e^+e^- collider and explain what detectors you would need. Why can this technique not be used in a hadron collider?

(3.8) This question is about the measurement of luminosity at a pp collider such as the LHC using Van der Meer (VDM) scans. Consider a collider with a revolution frequency f_{rev} and n_b colliding bunches. Let the numbers of particles in beams 1 and 2 be n_1 and n_2 and let the normalized bunch densities be $\rho_1(x, y)$ and $\rho_2(x, y)$. The luminosity is defined by the interaction rate for a process with a given cross section σ by $R = L\sigma$. In terms of the beam parameters, the luminosity is expressed by

$$L = n_b f_{\text{rev}} n_1 n_2 \int \rho_1(x, y) \rho_2(x, y) dx dy$$

Explain the origin of this equation. Assume that the particle densities are uncorrelated in the x and y directions. The luminosity can then be written as

$$L = n_b f_{\text{rev}} n_1 n_2 \Omega_x \Omega_y$$

where the beam overlap is defined in x (with an equivalent definition in y) as

$$\Omega_x = \int \rho_1(x) \rho_2(x) dx$$

Assume that the beams have Gaussian distributions with a common centre and root mean squares σ_{x1} and σ_{x2} . Show that

$$\Omega_x = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_{x1}^2 + \sigma_{x2}^2}}$$

Let $R(x)$ be the interaction rate as a function of the separation of the beams in the x direction:

$$R(x) \propto \exp \left[-\frac{x^2}{2(\sigma_{x1}^2 + \sigma_{x2}^2)} \right]$$

Therefore, the value of $\sigma_{x1}^2 + \sigma_{x2}^2$ can be determined from a VDM scan if the interaction rate for some particular process, $R(x)$, is measured as the separation between the two beams, x , is varied. Assuming

that the fraction of bunch crossings that register a hit in a counter is p , show that the mean number of collisions that produce such hits is given by $\mu = -\ln(1 - p)$. Why is it advantageous to use a small counter? What is the problem with this technique at very high luminosity? Suggest a possible detector technique that could be used for such a counter.

- (3.9) Synchrotrons have a periodic ring-shaped lattice of focusing and bending magnets. Specifying the position of a beam particle on the circumference of the accelerator by the distance s , the equations of motion in both horizontal and vertical directions about the ideal orbit have the form

$$\frac{d^2x}{ds^2} + K(s)x = 0$$

where $K(s + L) = K(s)$ is periodic. The solution is a quasiperiodic function of the form

$$x(s) = \sqrt{\varepsilon\beta(s)} \cos[\phi(s) - \phi_0]$$

where ε and ϕ_0 are constants. Both ε and $\beta(s)$ have dimensions of length. The phase $\phi(s)$ is related to the beta function $\beta(s)$ by $d\phi/ds = 1/\beta$. Explain the significance of the beta function.

Show that x and $x' \equiv dx/ds$ satisfy

$$\frac{x^2 + [\beta(s)x' + \alpha(s)x]^2}{\beta(s)} = \varepsilon$$

where

$$\alpha(s) = -\frac{1}{2} \frac{d\beta}{ds}$$

Reduce the expression to the standard form for an ellipse in (x, x') space and show that its area is $\pi\varepsilon$. This is the emittance of the beam (in one transverse direction). Explain its importance for accelerator design and control.

Particle detectors

4

4.1 Introduction

All the experimental discoveries that underpin our understanding of particle physics rely on particle detectors, so a good knowledge of how these sophisticated devices work is essential. The complexity of particle detectors has grown enormously from very simple beginnings to the very powerful detector systems used at the LHC. As in the rest of this book, we will not take a historical approach but try to find the easiest and most direct way to explain the fundamental physics. We will start in Section 4.2 with an overview of a collider detector, focusing on what the requirements are and giving a simple description of how the different subsystems are used to identify some types of particles and measure the energy of individual particles or ‘jets’. This will give us a good idea of what a collider detector looks like, but it will tell us nothing about how any particular detector actually works. To gain any useful understanding, we need to consider the basic detector physics that will explain quantitatively the performance of real detectors.

We start this systematic approach in Section 4.3 by considering how high-energy particles interact with matter and lose energy. The processes result in a relatively small number of electron–ion pairs, so the next issue to consider is how we can use this effect to create a measurable signal.¹ The fundamental detector physics of how signals are generated will be described in Section 4.4, since this step is obviously essential for any real understanding of how a particle detector works.

Armed with this knowledge, we can start to consider how basic particle detectors actually work. In Section 4.6, we will look at two techniques used for tracking the trajectory of charged particles: wire chambers and silicon detectors. Next, in Section 4.7, we will consider how to make energy measurements for charged and neutral particles in devices called calorimeters.

To select interesting events for permanent storage, while rejecting very high rates of background processes, very powerful trigger systems are required. We will review these briefly in Section 4.10, with a particular emphasis on LHC collider detectors, since these present the biggest challenges from the triggering perspective. Even with very powerful trigger systems, many petabytes of data are written to permanent storage every year at the LHC. Therefore, extremely powerful computer systems are required to process these data and to run the Monte Carlo simulation programmes used to understand the detector performance and correct

4.1 Introduction	71
4.2 Overview of collider detectors	72
4.3 Particle interactions with matter	74
4.4 Signal generation	81
4.5 Photon detection	85
4.6 Detectors for charged-particle tracks	86
4.7 Detectors for particle jets	98
4.8 Detectors for particle identification	105
4.9 Magnetic fields	106
4.10 Trigger	108
4.11 Examples of detector systems	109
Chapter summary	112
Further reading	112
Exercises	112

¹This is not strictly correct in silicon detectors, where we deal with electron–hole pairs.

²The biggest practical problem with very large computer farms is how to provide sufficient cooling to remove the heat.

for the inevitable imperfections. This computing requires $\sim 10^5$ CPUs, which would be difficult to deal with in one facility.² The problem has been solved by the use of GRID computing, in which the CPUs are distributed over many computer centres across the world. GRID computing is now a major research area in its own right, but will not be covered further in this book.

Having understood the basic building blocks, we will then look in Section 4.11 at how large particle detectors are designed and how they work. Here and in other chapters, we will use case studies of real detectors to see how the fundamental principles are applied in practice. Interestingly, we will see that there is no perfect solution to the many design challenges, and there are always difficult trade-offs in the design of any large detector. The discussion will focus on the design of the general-purpose LHC detectors, since these are the largest and most sophisticated detector systems ever built. We will also briefly consider neutrino detectors, since the constraints on these are not the same as those on collider detectors and the resulting systems are very different.

4.2 Overview of collider detectors

As an example of a collider detector, we will look at the general-purpose detectors at the LHC. As will be discussed in Chapter 13, the principle aims of the LHC are the study of the Higgs sector and the search for new physics beyond the Standard Model (SM), such as supersymmetry. Higgs bosons or any exotic particles will be heavy and will in general decay rapidly to SM particles, so we need to optimize the detector for energetic SM particles. We must be able to measure the momenta of photons, electrons, muons, taus, and hadron jets. As well as measuring the momenta, ideally we need to identify the different particles, which is non-trivial since the rates for hadron jets are $O(10^6)$ times higher than for leptons. We also need to distinguish between jets from b and c quarks and jets from light quarks. For neutrinos or exotic weakly interacting particles, such as SUSY WIMPs (see Section 13.4.1), direct detection is clearly impractical. However, we can infer the transverse momentum of these ‘invisible’ particles by using conservation of momentum. For this technique to be effective, we require detectors with calorimeters that cover most of the 4π solid angle (this technique is discussed in Chapter 13).

A very schematic view of the principle components of a general-purpose collider detector is shown in Fig. 4.1. Working our way out from the centre of the detector, we can see how the different elements contribute to satisfying these requirements:

- **Tracker:** This consists of very high-precision silicon detectors, immersed in a large magnetic field from a superconducting magnet. The trajectories of charged particles can be reconstructed, and hence the momenta can be computed. These detectors are used in

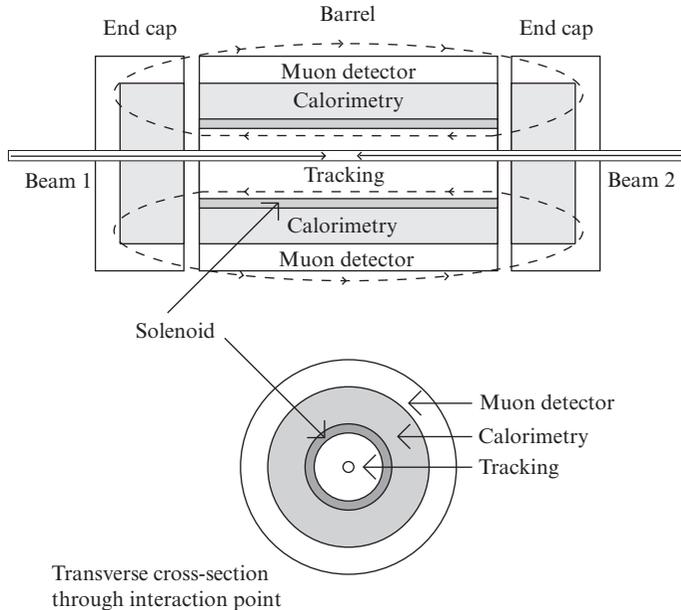


Fig. 4.1 Longitudinal and transverse views of a generic collider detector.

conjunction with the calorimeters and muon detector to identify and measure the momenta of electrons, muons, and taus. They can also measure the momenta of charged hadrons. The very high precision of these detectors allows good momentum resolution for high-momentum particles. It also allows b and c quarks to be identified, using the fact that because of their relatively long lifetime, the trajectories of their decay products do not point back to the primary vertex. The magnetic fields required, in the range 2–4 T, are created by superconducting magnets, which are based on similar technology to that used for the accelerator (see Chapter 3).

- Calorimeter:** The first aim of the calorimeters is to provide high-precision measurements of photons and electrons. The second aim is to measure the energy of hadrons and so reconstruct hadronic jets. All particles apart from muons and weakly interacting particles like neutrinos will deposit nearly all their energy in the calorimeters. In general, the energies are reconstructed from active detector elements interleaved with passive absorber material. For practical reasons that we will consider in Section 4.7, the calorimeters are divided into electromagnetic (EM) and hadronic sections. Each type of calorimeter is further divided into small cells, which enables reconstruction of the transverse and longitudinal profiles of the energy deposition. This provides very powerful separation between electrons, which deposit nearly all their energy in a small region of the EM calorimeter, and hadronic jets, which produce deeper and wider showers. To reconstruct the missing transverse momentum, it is essential that the calorimeter cover a solid angle as near to 4π as possible.³

³Two holes around the beam pipes are unavoidable, so there can be significant energy ‘lost’ down the pipes, but as the angles are very small, the transverse momenta are low; hence we can measure missing transverse momentum but not missing longitudinal momentum.

- **Muon spectrometer:** If the calorimeter is sufficiently thick, the majority of particles emerging from the calorimeters will be muons (ignoring neutrinos), because they do not tend to produce electromagnetic showers like electrons or have hadronic interactions. The trajectories of the muons are measured in large wire chambers and can be matched to high-transverse-momentum charged particles measured in the tracker, reducing the effects of hadrons ‘leaking’ out of the back of the calorimeter. If there is a magnetic field in the region of the muon chambers, the muon trajectory can be used to determine the muon momenta. Possible magnetic field configurations are considered in Section 4.9. The momenta of the muons can be measured independently in the tracker and combined with the measurement in the muon spectrometer to get the best precision.

4.3 Particle interactions with matter

This section covers the most important interactions of high-energy particles with matter that are needed to understand detector physics. For tracking detectors, the most important process is ionization, since this generates the electron–ion pairs that we can detect. Multiple scattering is also important in tracking detectors because it limits the resolution. Electromagnetic processes such as pair production are fundamental for understanding electromagnetic and hadronic calorimeters. Finally, hadronic interactions are obviously of particular importance for understanding hadronic calorimeters.⁴

⁴This is of course a very simplified picture—in reality, all these processes affect all detector types to some extent.

4.3.1 Ionization

All charged particles interact with electrons in the atoms in any material in the detector. For high-energy particles,⁵ the energy transferred to the electrons can be larger than the ionization energy, thus creating free electrons and positive ions. How to detect such secondary charged particles is discussed in this chapter. These collisions result in the incident particle losing energy in the lab frame (they are approximately elastic collisions in the CMS).⁶ We can understand the main features of ionization energy loss by starting from the formula for Rutherford scattering. The differential cross section (see Exercise 4.1) as a function of the 4-momentum transfer Q^2 and speed of the incoming particle β is given by

$$\frac{d\sigma}{dq^2} = \frac{4\pi\alpha^2}{q^4\beta^2} \quad (4.1)$$

where z is the charge in units of electron charge of the target particle by which the charged particle is scattered and α is the fine-structure constant. We can evaluate Q^2 in the rest frame of the electron before

⁵Here we are typically interested in particles with energies $E \gg 1$ MeV.

⁶Charged particles can also lose energy by interacting with the atomic nuclei, but the energy transferred by the elastic scattering considered in this section is negligible compared with interactions with electrons because of the larger mass of the nuclei compared with electrons. Electrons also lose energy by bremsstrahlung, and this will be discussed in Section 4.3.3.

the collision to be (see Exercise 4.2) $Q^2 = 2m_e T$, where T is the kinetic energy of the scattered electron. Then, changing variables in eqn 4.1,

$$\frac{d\sigma}{dT} = \frac{2\pi z^2 \alpha^2}{m_e \beta^2 T^2} \quad (4.2)$$

We then convert this expression for the energy loss in one collision to the average energy loss as a charged particle interacts with many atoms in some medium. The rate of energy loss per unit length in a medium with N atoms per unit volume and atomic number Z is

$$\frac{dE}{dx} = NZ \int_{T_{\min}}^{T_{\max}} T \frac{d\sigma}{dT} dT \quad (4.3)$$

The minimum energy T_{\min} is related to the ionization energy I . We can calculate the maximum kinetic energy of the electron in the lab frame by considering a collision in the rest frame in which the direction of motion of the electron is reversed (see Exercise 4.2), which gives $T_{\max} = 2\beta^2 \gamma^2 m_e$. Substitution into eqn 4.3 gives an approximate formula for the rate of energy loss of charged particles:

$$\frac{dE}{dx} = \frac{2\pi NZ z^2 \alpha^2}{m_e \beta^2} \ln \left(\frac{2\gamma^2 \beta^2 m_e}{I} \right) \quad (4.4)$$

This shows that the energy loss initially decreases with increasing energy and then rises logarithmically with energy. Finally, this formula modified by relativistic effects is known as the Bethe formula⁷

$$\frac{dE}{dx} = K \frac{Z}{A\beta^2} \left[\ln \left(\frac{2m_e \beta^2 \gamma^2}{I} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \quad (4.5)$$

where $K = 4\pi N_A r_e^2 m_e$ (N_A is Avogadro's number and r_e is the classical radius of the electron) and Z and A are the atomic number and atomic mass number of the nucleus. It is conventional to express the stopping power in units of $\text{MeV g}^{-1} \text{cm}^2$. To transform this to the stopping power per unit length, we simply multiply by the density ρ . At relativistic energies, the electric field from the primary charged particle flattens and so allows collisions with more distant atoms. However, at very high energy, this effect is reduced by the polarization of the medium, which leads to the 'density effect' correction term $\delta(\beta\gamma)$.

The mean energy loss for charged particles in different media as a function of $\beta\gamma$ is shown in Fig. 4.2. The important features of the stopping power are very similar for all targets: at low momentum, the stopping power decreases rapidly as the momentum of the incident particle increases, and then rises logarithmically at higher momentum. There is a broad minimum around $\beta\gamma \sim 3$ and the value of the minimum is typically in the range $1\text{--}3 \text{ MeV g}^{-1} \text{cm}^2$. Note that the energy loss by ionization scales with the Z of the material, which is very different to the Z^2 scaling that we find for pair-production and bremsstrahlung processes. We have discussed the mean energy loss, but there can be very

⁷This formula used to be called the Bethe–Bloch formula, but it is now usually called the Bethe formula.

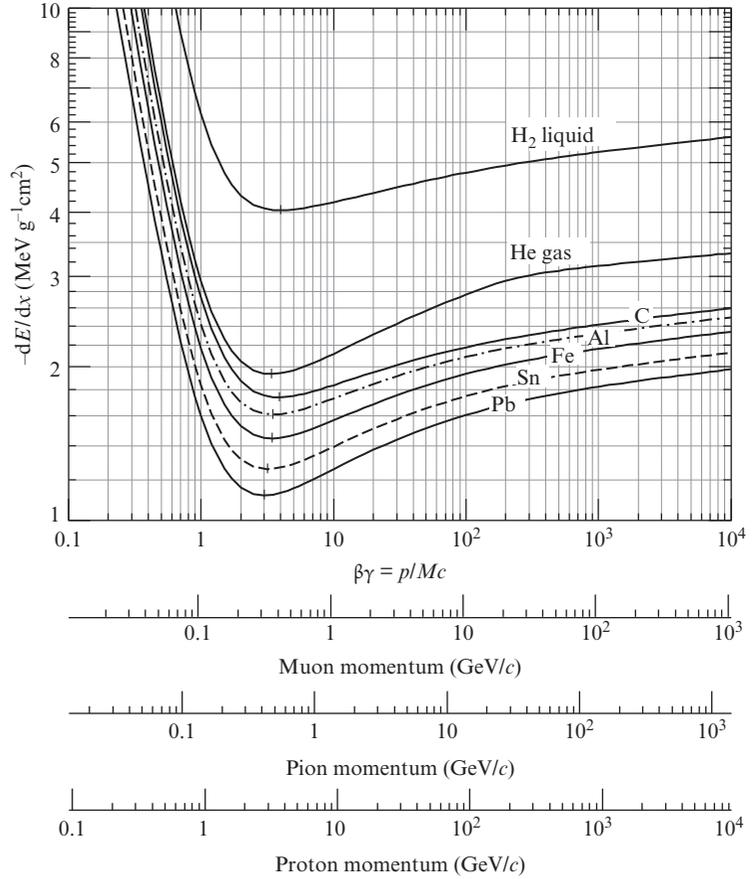


Fig. 4.2 The mean energy loss by ionization in different materials as a function of $\beta\gamma$ [115] (where β and γ are the usual relativistic factors). Note the units of $\text{MeV g}^{-1} \text{cm}^2$. To convert to linear stopping power, multiply by the density.

⁸Here the effect of scattering off the atomic nucleus dominates over that from atomic electrons because of the larger charge on the nucleus.

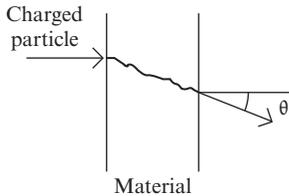


Fig. 4.3 A charged particle undergoing multiple scattering in a material is deflected by an angle θ .

large fluctuations because of the large range of energies that can be lost in a single collision. The spread in the actual energy lost is given by the very broad ‘Landau’ distribution. Very large ‘tails’ in the distribution are caused by the emission of single relatively energetic electrons (called ‘ δ -rays’).

4.3.2 Multiple scattering

When a charged particle traverses a slab of a material, as sketched in Fig. 4.3, it undergoes a very large number of very small-angle Coulomb scatters from the nuclei of the material.⁸ The net result of that is that the particle emerges from the slab at an angle θ with respect to the initial direction. Considering many identical particles, one gets a distribution of their angles θ (in a plane like the plane of Fig. 4.3, or for any plane containing the initial direction vector) that follows a Gaussian distribution with standard deviation

$$\theta_0 = \frac{13.6 \text{ MeV}}{\beta p} q \frac{x}{X_0} \left(1 + 0.038 \ln \sqrt{\frac{x}{X_0}} \right) \quad (4.6)$$

where X_0 is the radiation length (see Section 4.3.3), x is the thickness of the material slab, and p , β , and q are respectively the momentum, velocity, and charge of the particle. The root mean square displacement of the particle trajectory y is then $y_{\text{plane}}^{\text{rms}} = (x/\sqrt{3})\theta_0$. The effects of multiple scattering degrade the resolution of track reconstruction and therefore can have profound effects on the detector performance, as will be discussed later in this chapter. Note that the amount of multiple scattering scales with the amount of material expressed in radiation lengths. This provides one motivation for the design of tracking detectors that are ‘thin’ in units of radiation length. As the radiation length scales with Z^{-2} , this shows that we should minimize the use of high- Z material.

4.3.3 Electromagnetic interactions

Electrons and positrons lose energy by ionization in a similar way to that discussed in Section 4.3.1 (although there are some differences, associated with issues like the spin and identical particles for the case of electrons). However, at high energy, the dominant process for energy loss is bremsstrahlung (see Fig. 4.4). For an electron of energy E , the rate of change of energy due to bremsstrahlung as a function of distance x is given by

$$\frac{dE}{dx} = -\frac{E}{X_0} \quad (4.7)$$

where X_0 is the radiation length for the material. We can easily integrate eqn 4.7 to show that in travelling a distance X_0 , the electron energy decreases by a factor of $1/e$. An approximate formula for the radiation length is given by (see [115] for the full expression)

$$\frac{1}{X_0} \sim \frac{4\alpha^3}{m_e^2} \frac{N_A}{A} Z^2 L_{\text{rad}} \quad (4.8)$$

N_A is Avogadro’s number, Z is the atomic number and A is the atomic mass number (number of protons and neutrons). where for $Z > 4$, $L_{\text{rad}} = \ln(184.15Z^{-1/3})$. We can see that the radiation length scales as $1/\alpha^3$ as expected because the Feynman diagram contains three vertices. The electron ‘sees’ the charge of the entire nucleus at one vertex, so the cross section scales with the atomic number as⁹ Z^2 . The differential cross section for bremsstrahlung as a function of the variable $y = k/E$, where k and E are respectively the photon and electron energies, is given to a good approximation by

$$\frac{d\sigma}{dk} = \frac{A}{X_0 N_A k} \left(\frac{4}{3} - \frac{4}{3}y + y^2 \right) \quad (4.9)$$

The characteristic feature of eqn 4.9 is that the photon energy spectrum is peaked at low values. In one radiation length, it is very unlikely that the electron will lose energy to only one high-energy photon—it is far more common for it to lose energy to many lower-energy photons.

High-energy photons can undergo pair conversion (see Fig. 4.5), which is clearly a process closely related to bremsstrahlung. At high energies,

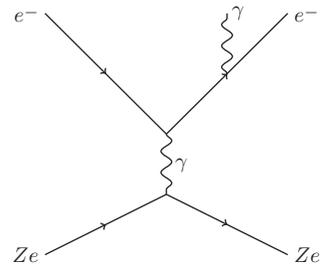


Fig. 4.4 Lowest-order Feynman diagram for bremsstrahlung ($eZ \rightarrow eZ\gamma$) in a material with nuclear charge Ze .

⁹This scaling with Z is more rapid than the linear scaling with Z that we found for energy loss by ionization.

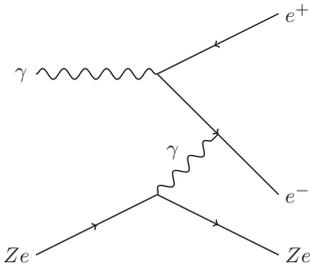


Fig. 4.5 Lowest-order Feynman diagram for the pair production process for a photon interacting with a nucleus of charge Ze .

¹⁰Note that the length for a primary electron to decrease in energy by a factor f is $\frac{7}{9}$ of the length for which the probability of a photon not to pair-produce is equal to the same factor f .

the differential cross section for pair production as a function of the fraction of the photon energy given to the electron, x , is

$$\frac{d\sigma}{dx} = \frac{A}{X_0 N_A} \left[1 - \frac{4}{3}x(1-x) \right] \quad (4.10)$$

We can integrate eqn 4.10 to obtain the total pair production cross section¹⁰

$$\sigma = \frac{7}{9} \frac{A}{X_0 N_A} \quad (4.11)$$

At lower energies, the dominant process for energy loss by photons is Compton scattering $\gamma e \rightarrow \gamma e$. Here the incident electron is approximately at rest in an atom and it is ejected from the atom in the process (i.e., in the lab frame, energy is transferred from the incident photon to the outgoing electron).

Now that we have considered the fundamental electromagnetic interactions in matter, we can understand the nature of the resulting electromagnetic showers. Incident high-energy electrons will lose energy by bremsstrahlung and the resulting photons will create e^+e^- pairs, which in turn will create more photons by bremsstrahlung. We need to consider the competition between the rate of energy loss from bremsstrahlung/pair production and ionization. The former increases approximately linearly with energy, whereas loss due to ionization increases only logarithmically. When the energy of the electrons decreases to the ‘critical energy’ E_c , the energy loss by bremsstrahlung will be equal to that by ionization. An approximate fit to the critical energy as a function of atomic number Z is given by

$$E_c = \frac{610}{Z + 1.24} \text{ MeV} \quad (4.12)$$

¹¹We can make an average correction to allow for shower leakage out of the back of the calorimeter, but there will always be shower-to-shower statistical fluctuations in the amount of leakage, which we cannot correct for. So, if we want a high-resolution electromagnetic calorimeter, we must ensure that it is deep enough for almost complete shower containment.

¹²This sets the natural size for the transverse granularity for electromagnetic calorimeters. We wish to separate electrons (positrons) from hadrons using, among other measures, the transverse shower size. This improves with finer granularity, but we clearly do not gain by having cells with lateral dimensions much less than R_M .

As the electron (positron) energies become lower than E_c , they lose energy rapidly, become non-relativistic and lose energy by ionization even more rapidly, hence ending the shower development. This results in the shower depth varying logarithmically with energy (see Exercise 4.4). The longitudinal shower profile can be calculated rather accurately using Monte Carlo simulations and an example is shown in Fig. 4.6. We require nearly complete shower containment to obtain good energy resolution, so for 30 GeV electrons we need a depth of at least $\sim 20X_0$.¹¹

Electromagnetic showers broaden as they penetrate deeper into matter owing to multiple Coulomb scattering of the electrons (positrons) and the scattering angles involved in bremsstrahlung and pair production. The first effect dominates and we can parameterize the width of the shower by the ‘Molière radius’ $R_M = X_0 E_s / E_c$, with $E_s \approx 21$ MeV. Approximately 90% of the energy is contained within a radius of R_M .¹²

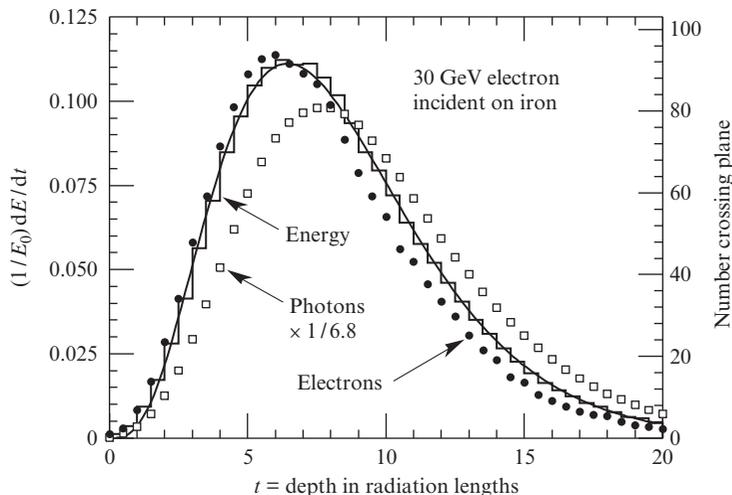


Fig. 4.6 Simulation of the longitudinal shower profile for incident 30 GeV electrons and photons on iron [115]. The histogram shows the energy deposition and the circles (squares) indicate the number of electrons (photons). The photons penetrate more deeply than electrons, reflecting the factor of $\frac{7}{9}$ in eqn 4.11.

4.3.4 Čerenkov radiation

When a charged particle moves at a speed v greater than the local phase velocity of light, $1/n$, where n is the refractive index of the medium, it will emit Čerenkov photons. The angle of the Čerenkov photons relative to the charged particle can be calculated using a simple geometrical argument (see Fig. 4.7). In time t , the relativistic particle travels from A to B, a distance of vt . The electromagnetic wave emitted by the particle from A is travelling at the (lower) speed of $1/n$. The wavefront, defined by the plane with a constant phase, is given by the line from C to B. Hence the Čerenkov angle is given by $\cos \theta_C = 1/(nv)$.¹³ The photons are typically in the optical range and can be detected in a similar way to that used for scintillation light (see Section 4.4.2).

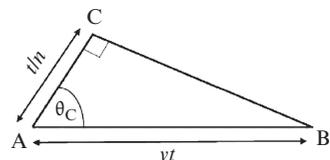


Fig. 4.7 Geometrical construction for calculation of the Čerenkov angle.

¹³We are using natural units with $c = 1$ and we have assumed that the medium is non-dispersive.

4.3.5 Transition radiation

If a high-energy charged particle crosses a boundary between two media with different dielectric constants, it can emit transition-radiation photons. The yield depends on the Lorentz factor γ and therefore allows the separation of electrons from charged hadrons. The yield per interface is $O(\alpha)$ and is therefore very low, implying that a practical transition-radiation detector requires hundreds of interfaces, which can be achieved for example with Mylar foils.

4.3.6 Hadronic interactions

High-energy hadrons undergo nuclear interactions in matter. The physics involved cannot be calculated from first principles, so phenomenological models are needed. It is useful to define the *interaction length* λ_I as the length in a material in which the probability of a hadron not interacting

Element	X_0 (g cm ⁻²)	λ_I (g cm ⁻²)	ρ (g cm ⁻³)
Iron	13.84	132.1	7.874
Copper	12.86	137.3	8.960
Lead	6.37	199.6	11.350
Uranium	6.00	209.0	18.950

Table 4.1 Radiation length X_0 , interaction length λ_I , and density ρ for some elements.

¹⁴This provides another motivation for using high- Z absorbers in an electromagnetic calorimeter (apart from wishing to limit the depth required for good shower containment): electromagnetic showers are contained in a shorter depth than hadronic showers and the separation is better for higher- Z absorbers.

is $1/e$. The cross section at high energy for scattering of a hadron on a nucleus scales like $\sigma = R_0 A^{2/3}$, which is quite different to the Z^2 scaling for bremsstrahlung and pair production.¹⁴ The interaction length is compared with the radiation length for a few common absorbers used in calorimeters in Table 4.1. The longitudinal shower profile for high-energy pions in iron [88] is shown in Fig. 4.8. We see that for good shower containment, we need a depth of about $10\lambda_I$, which results in very large calorimeters. This obviously increases the cost of the hadronic calorimeter itself, but also increases the radius for the start of the muon detectors, and thus increases the area and cost of the muon spectrometer. Therefore, for cost reasons, there will usually be some significant energy leakage out of the back of a hadronic calorimeter. As for electromagnetic calorimeters, we can make an average correction for this effect, but the statistical fluctuations will degrade the resolution.

A high-energy hadron interacting with a nucleus will create a mixture of charged and neutral hadrons. The π^0 s will decay rapidly to photons

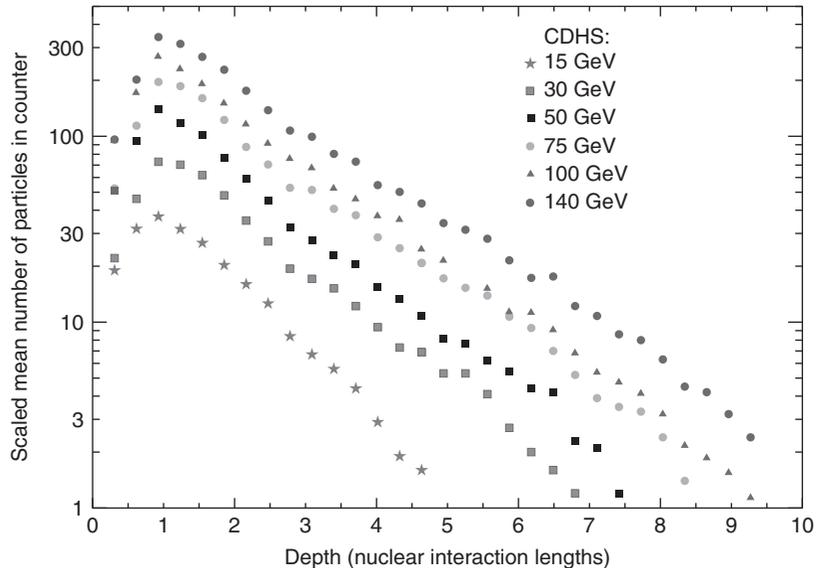


Fig. 4.8 Measured shower profiles for high-energy pions in the CDHS detector [88].

and thus induce electromagnetic showers. The charged hadrons produced will penetrate further into the calorimeter and create secondary hadronic interactions, leading to the development of hadronic showers deep into the calorimeter. The big difficulty with hadronic calorimetry is that a significant fraction of the energy goes into nuclear breakup and evaporating neutrons and protons from the nuclei. The resulting low-energy nuclei and protons will be very heavily ionizing and lose energy rapidly. Typically, this will occur in the passive absorber,¹⁵ producing no detectable signal in the ‘active’ layers. The low-energy neutrons will scatter and thermalize on a timescale of microseconds, and so any photons produced from neutron capture will be outside the time ‘window’ for signal collection. The fraction of energy that is effectively ‘lost’ in a hadronic interaction due to these processes is very large (typically in the range 20–40%). The real problem is that there is a very large variation in this lost fraction from shower to shower, which greatly degrades the resolution of hadronic compared with electromagnetic calorimeters. The magnitude of the effect can be parameterized by the ratio of the response to electrons to that to hadrons, e/h . If e/h is significantly different from unity, the calorimeter resolution will be limited and there will be large non-Gaussian fluctuations. Several ideas have been pursued to try to achieve ‘compensating’ calorimeters in which $e/h \approx 1$ and these will be discussed in Section 4.7.5.

¹⁵For cost reasons, hadronic calorimeters are divided into alternating layers of ‘passive’ absorber and ‘active’ layers that detect the signal.

4.4 Signal generation

In Section 4.3, we have considered how particles lose energy in matter and create showers of secondary charged and neutral particles. We now need to examine how we can actually detect these secondary particles as well as the particles from the primary collision. In Section 4.4.1, we will see how charged particles moving between electrodes induce currents, which we can amplify and read out with suitable electronics. Another approach, considered in Section 4.4.2, is to use scintillation light. The scintillation and Čerenkov processes result in photons in the visible or ultraviolet wavelengths, so in Section 4.5 we review techniques to detect these photons.

4.4.1 Moving charges

In this section, we explain how to calculate the induced currents created by moving charges, which generate the electrical signals we can measure in detectors like wire chambers or silicon detectors. We first calculate the induced current for a simple case and then discuss the general solution.

Consider a charged particle held between the two (infinite) plates of a parallel-plate capacitor, with both plates grounded. The potential is given by the solution of Laplace’s equation [76], subject to appropriate

boundary conditions (the potential is 0 on the plates and approximates that from a point charge in the vicinity of the charge):

$$V(\rho, z) = \frac{q}{\epsilon_0 \pi L} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi z}{L}\right) \sin\left(\frac{n\pi z_0}{L}\right) K_0\left(\frac{n\pi \rho}{L}\right) \quad (4.13)$$

where z_0 is the distance from the lower plate to the point charge, L is the separation between the plates, $\rho = \sqrt{x^2 + y^2}$ (where x and y are the Cartesian coordinates of the point charge in the plane of the lower plate—see Fig. 4.9), and K_0 is a modified Bessel function. The solutions for three locations of the charge are illustrated by the equipotentials shown in Fig. 4.9.

The induced electric surface charge density on the conducting plate at $z = 0$ is given by $\sigma = \epsilon_0 |E_z(z = 0)|$, where $\mathbf{E} = -\nabla V$ is the electric field evaluated at the edge of the conductor (z is the direction perpendicular to the conductor). When the charge is near the upper (lower) plate, we see that the equipotentials are more tightly packed near the upper (lower) plate. Therefore, when the charge is near the upper (lower) plate, the E field will be larger nearer the upper (lower) plate and hence there will be a larger induced charge on the upper (lower) plate. The fields and induced charges are obviously symmetric when the charge is equidistant from the two plates. Now let us imagine moving the charge from near the upper plate to near the lower plate. Initially, most of the induced charge will be on the upper plate, but this will gradually change and at the end most of the induced charge will be on the lower plate. This then looks like a current flowing between the two conductor plates. This is a qualitative example of the fundamental result in detector physics; moving charges

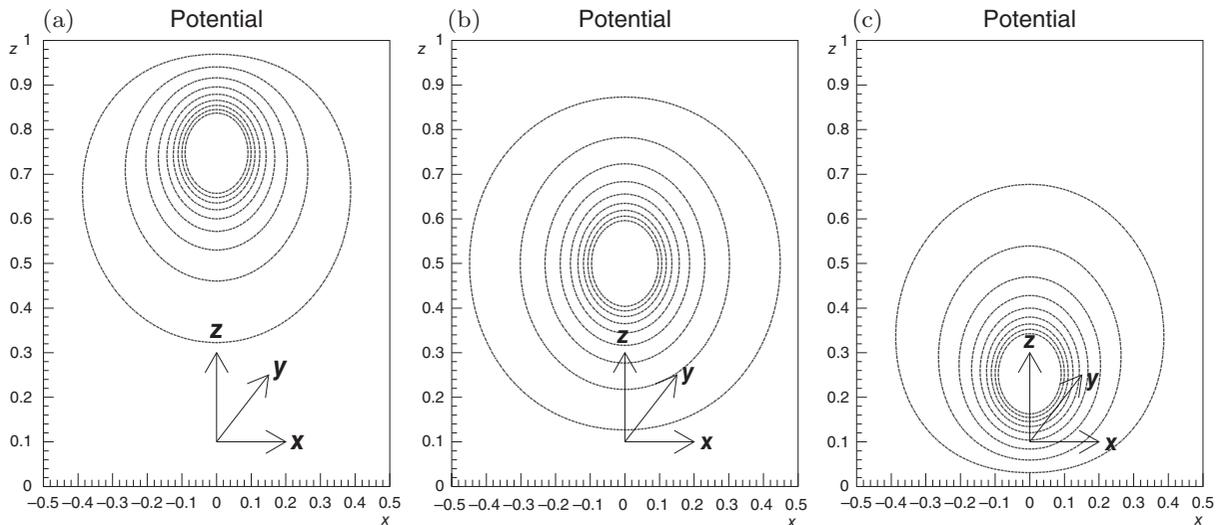


Fig. 4.9 Equipotentials (arbitrary units) for a point charge at three different locations in a parallel-plate capacitor: (a) near the upper plate; (b) near the centre; (c) near the lower plate. The equipotentials near the point charge are omitted for clarity.

between conducting electrodes induce currents.¹⁶ This current can be amplified and digitized by appropriate readout electronics.

Now that we have seen a qualitative description of the physics of induced charges, we can look at the quantitative solution. Taking the derivative in the z direction of the potential (eqn 4.13), we can determine the induced surface charge density on the upper and lower plates using Gauss's law:

$$\sigma(\rho, z = 0) = -\frac{q}{(L\pi)} \sum_{n=1}^{\infty} \frac{n\pi}{L} \sin\left(\frac{n\pi z_0}{L}\right) K_0\left(\frac{n\pi\rho}{L}\right) \quad (4.14)$$

$$\sigma(\rho, z = L) = \frac{q}{(L\pi)} \sum_{n=1}^{\infty} \frac{n\pi}{L} (-1)^n \sin\left(\frac{n\pi z_0}{L}\right) K_0\left(\frac{n\pi\rho}{L}\right)$$

We can integrate the surface charge density to find the total charge induced on the upper plate as

$$\begin{aligned} Q_U &= 2\pi \int_0^{\infty} \sigma(\rho, z = L) \rho \, d\rho \\ &= \frac{2q}{L} \sum_{n=1}^{\infty} \frac{L}{n\pi} (-1)^n \sin\left(\frac{n\pi z_0}{L}\right) \int_0^{\infty} x K_0(x) \, dx \end{aligned} \quad (4.15)$$

The integral is equal to unity, so

$$Q_U = \frac{2q}{L} \sum_{n=1}^{\infty} \frac{L}{n\pi} (-1)^n \sin\left(\frac{n\pi z_0}{L}\right) \quad (4.16)$$

The infinite sum is related to a Fourier series (see Exercise 4.3), so

$$Q_U = -\frac{qz_0}{L} \quad (4.17)$$

We can calculate the surface charge on the lower plate at $z = 0$ by the same method, to obtain¹⁷

$$Q_L = -\frac{q(L - z_0)}{L} \quad (4.18)$$

Now let the charge between the capacitor plates move with a speed v in the negative z direction. The induced charge flows from the upper to the lower plate and the current (while the charge is moving) is given by the rate of change of charge as

$$I = -\frac{qv}{L} \quad (4.19)$$

We have determined the induced current for the simplest possible geometry.

A more general solution to the calculation of the induced current, which is indispensable for understanding realistic detector geometries,

¹⁶Note that the induced signal occurs as long as the charge is moving and stops when the charges are collected on the electrodes. A popular misconception is that the signal only arises when the charge is 'collected' at an electrode.

¹⁷The total induced charge on the two plates is $-q$, as expected.

¹⁸See Spieler in the Further Reading at the end of this chapter for a derivation of Ramo's theorem.

¹⁹Note that this field is not the same as the electric field and does not even have the same dimensions.

is provided by Ramo's theorem.¹⁸ This will provide us with a simple method for calculating the induced currents from any movements of charges and is therefore of fundamental importance in detector physics. First, we set the potential on the electrode being considered to 1 V and apply 0 V to all other electrodes and calculate the potential Φ by solving Laplace's equation subject to these boundary conditions. The 'weighting' field¹⁹ is defined as

$$\mathbf{E}_W = -\nabla\Phi \quad (4.20)$$

The current induced on this electrode, caused by the motion of n charges q_j , moving with velocities \mathbf{v}_j ($j = 1, \dots, n$), is given by

$$i = -\sum_{j=1}^n q_j \mathbf{v}_j \cdot \mathbf{E}_W \quad (4.21)$$

²⁰Only in the case of two-electrode systems does the weighting field have the same form as the physical electric field.

The velocity depends on the real electric field, not the weighting field.²⁰ As a simple 'sanity check', we can now use Ramo's theorem to calculate the induced current for the case of a point charge between the plates of an infinite parallel-plate capacitor and compare it with the result obtained above. For this geometry, if we apply 1 V on one electrode and 0 V on the other electrode, then the weighting field is uniform and has a magnitude of $1/L$. For a point charge q moving with velocity v parallel to this weighting field, we obtain the induced current from eqn 4.21 as

$$I = -\frac{qv}{L} \quad (4.22)$$

²¹This result is independent of the forces causing the charge to move with a velocity v . In a particle detector, the motion is due to the applied electric and magnetic fields and the interactions of the moving charge with atoms or molecules in the detector.

which is in agreement with eqn 4.19.²¹

4.4.2 Scintillators

Scintillators are materials in which ionizing particles can cause scintillation light, which can be detected by photodetectors. There are two broad classes of scintillator: organic and inorganic. A common example of an organic scintillator is polystyrene. In an organic scintillator, molecules are lifted into an excited state by an ionizing particle and de-excite by emitting scintillation photons (typically in the UV). The problem with this is that the reverse reaction has a large cross section, so these UV photons will be rapidly re-absorbed.²² This problem is solved by introducing a dopant so that these photons are absorbed by a fluorescent molecule (a 'fluor'). The fluor then decays rapidly to a lower-energy state via a radiative decay, emitting longer-wavelength photons. This increases the attenuation length, but it is usually still too short for practical applications. Therefore, a secondary fluor is used to shift the photons into the visible wavelength range, and these photons can have a suitably long attenuation length. The typical scintillation and fluorescence processes [115] are illustrated in Fig. 4.10. This type of organic scintillator is often used in sampling calorimeters (see Section 4.7).

²²This is clearly a problem for an application requiring large-area scintillators.

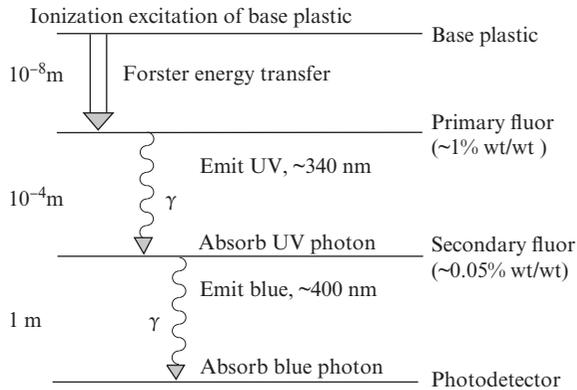


Fig. 4.10 Scintillation and fluorescence steps in an organic scintillator. [115]. Typical values are given for the wavelength and absorption length of the photons. The first step in the process, ‘Foerster energy transfer’, does not involve photon emission but is a dipole–dipole interaction between the base and the primary fluor.

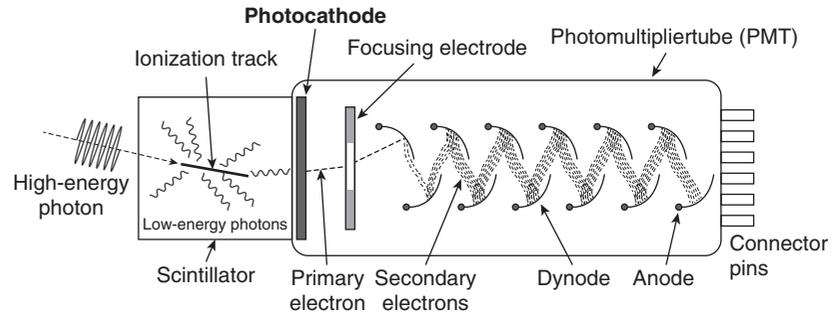
A classic example of an inorganic scintillator is thallium-doped sodium iodide, NaI(Tl). A high-energy particle can excite an electron from the valence to the conduction band. The electron can drop from the conduction to the valence level with the emission of a photon. However, the reverse process will result in too short an attenuation length for a useful detector. Therefore, a different process is used in which high-energy particles create excitons (loosely bound states of an electron and a hole). An exciton can move through the crystal until it is captured by an impurity state (created by the doping with Tl), which can then decay via emission of a photon, thus creating scintillation light.²³ This has the advantage of high density, which allows the construction of a more compact calorimeter, thus reducing the cost, and it also has a very good yield for scintillation light. This scintillator is still used in many applications and it was used in older particle physics detectors. The problem is that it is too slow for use at modern colliders, because the scintillation decay time is ≈ 250 ns, which is much longer than the time between collisions at the LHC of 25 ns. To use an inorganic scintillator at the LHC, we need a very fast decay time. Also, the scintillator must be very tolerant to radiation—most scintillators would become opaque after exposure to LHC radiation levels. Such a scintillator, PbWO_4 , has been developed for the CMS electromagnetic calorimeter; its use there will be described in Section 4.7.2.

²³As the doping concentration is relatively low, the probability of the scintillation light being reabsorbed in the crystal is very low; i.e. the crystal is transparent at this wavelength.

4.5 Photon detection

We have seen that scintillation and Čerenkov radiation result in photons in the range from the optical to the UV, which we have to convert into an electrical signal that can be digitized and read out. The traditional method is based on photomultipliers (PMTs), but another technique that is becoming increasingly common uses avalanche photodiodes. A schematic illustration of a photomultiplier coupled to a scintillator is shown in Fig. 4.11. A photomultiplier has a photocathode, usually

Fig. 4.11 Schematic view of a photomultiplier and the main processes involved. The primary photon is emitted from the photocathode and is accelerated and focused until it hits the first dynode. It then liberates many secondary electrons, which are accelerated to the next dynode. The resulting induced current is detected on the anode. From <https://commons.wikimedia.org/wiki/File:PhotoMultiplierTubeAndScintillator.jpg>



containing two alkali elements to obtain the best quantum efficiency. When a photon with an energy greater than the work function hits the photocathode, it can emit an electron by the photoelectric effect.²⁴ The resulting electron is then accelerated by an applied electric field until it hits the first dynode.²⁵ This causes the emission of several secondary electrons (the electron has been accelerated so it has sufficient energy to do this). The secondary electrons are similarly accelerated and strike the second dynode. This clearly multiplies the number of electrons (hence the name photomultiplier). Several stages of dynodes are used and it is easy to obtain a very large gain ($\sim 10^6$ or more). A single photon thus creates a large pulse of electrons that is easy to detect and digitize. One disadvantage of PMTs is that they do not work in large magnetic fields.²⁶

The simplest solid state photodetector is a photodiode. In a photodiode, photons create electron-hole pairs in a detector working on the same principles as that of a silicon detector (see Section 4.6.2).²⁷ The problem is that the small signal results in a low signal-to-noise ratio. In an avalanche photodiode (APD), the electric field is large enough that electrons acquire sufficient energy to create further electron-hole pairs, leading to an 'avalanche' effect. This avalanche process creates an intrinsic gain in the device that results in APDs having better resolution for small calorimeter signals than simple photodiodes. This requires that a larger reverse bias be applied, typically ~ 100 V, which results in an avalanche gain in the range 10–100. The gain of an APD is more sensitive to the applied bias voltage and the temperature than that of a simple photodiode. In addition, the design needs to ensure that the avalanche does not lead to electrical breakdown. One key advantage of APDs for particle physics applications is that they are insensitive to applied magnetic fields.

4.6 Detectors for charged-particle tracks

The traditional technology for tracking used wire chambers. These have been largely replaced by silicon detectors for the inner 'trackers' in LHC general purpose detectors. However, the cost of silicon detectors would

²⁴This is called a photoelectron.

²⁵Additional electrodes act as electrostatic focusing elements to increase the fraction of electrons collected at the first dynode.

²⁶For operation in moderate magnetic fields, PMTs can be shielded by shields made of 'mu-metal', an alloy with an exceptionally large relative permeability. However, saturation effects prevent this technique from working in high magnetic fields.

²⁷Silicon is one possible semiconductor that can be used for photodiodes, but there are photodiodes made from other semiconductors such as GaAs or In-GaAs. The optimal choice for any application depends on several factors, including wavelength, speed of response, and cost.

be prohibitive for the very large-area outer detectors needed for the muon spectrometers, so wire chambers are the only practical technology for these systems.

4.6.1 Wire chambers

A primary high-energy charged particle passing through a gas will create a few electron–ion pairs by ionization. To create a sufficiently large signal (i.e. greater than the electronic noise of an amplifier), we need to use an avalanche process. We start by considering the ‘gas gain’ caused by an avalanche, then consider the simple proportional wire chamber, and finally look at a ‘drift’ chamber.

Gas gain

At sufficiently high electric fields ($\sim 100 \text{ kV cm}^{-1}$), electrons drifting in an electric field acquire sufficient energy to cause further ionization in the gas and thus enable an avalanche process that can result in a very large increase in the number of electron–ion pairs. We define the gas gain $G = N/N_0$, where N_0 and N are the initial and final numbers of electron–ion pairs. The change in N with a distance travelled ds is

$$dN = N\alpha ds \quad (4.23)$$

where α is called the first Townsend coefficient and has to be measured experimentally. We can integrate eqn 4.23 for the gas gain:

$$\begin{aligned} G = N/N_0 &= \exp\left(\int \alpha ds\right) \\ &= \exp\left(\int_{E_{\min}}^{E_{\max}} \frac{\alpha}{dE/ds} dE\right) \end{aligned} \quad (4.24)$$

where E is the electric field, E_{\min} is the value of E at the start of the avalanche, and E_{\max} is the value at the end of the avalanche (e.g. at the wire in a wire chamber). The value of E_{\min} is simply related to the mean free path for electrons λ and the average ionization energy I by conservation of energy: $eE_{\min}\lambda = I$.

We can now summarize the general features of the gas gain as a function of the applied voltage across a chamber as illustrated in Fig. 4.12. At very low voltages, the electrons recombine with ions before they are collected. At higher voltages, we can distinguish different regions:

- **Ionization chamber:** In this region, the electrons do not acquire sufficient energy to start an avalanche and we just see the signal from the primary electron–ion pairs, i.e. there is no gas gain.
- **Proportional regime:** In the ‘proportional’ regime, there is gas gain and the number of electron–ion pairs created by the avalanche is proportional to the number of primary electron–ion pairs. Typical values of gas gain are in the range 10^4 – 10^5 .

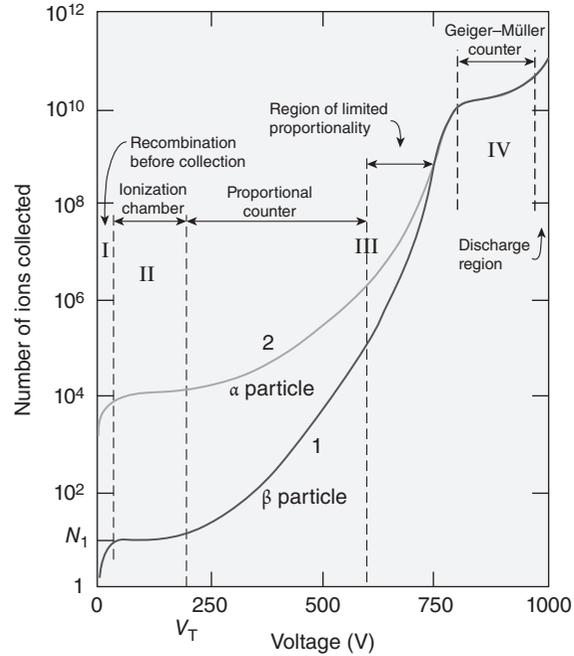


Fig. 4.12 Variation in gas gain as a function of applied voltage [125].

- **Limited proportionality:** At higher voltages, the density of the electron-ion pairs is so high that after the lighter electrons have drifted some distance, the net space-charge density is so large that it decreases the field, thus lowering the gain.
- **Geiger-Müller mode:** The highest voltages result in the Geiger-Müller mode with very large gains $\sim 10^{10}$. For operation in this regime, to avoid complete electrical breakdown, we require a very large series resistance for the high voltage to lower the latter when the current gets too high. This creates a long recovery time between pulses and is therefore not useful for high-rate applications.

The actual calculation of the gas gain depends on dE/ds , which clearly depends on the geometry used to create the field. For example, we can calculate the gas gain for the case of the proportional wire chamber to be discussed in detail below. Substituting for the electric field from eqn 4.27 into the gas gain equation (eqn 4.24), we can show that the gas gain is

$$G = \exp \left\{ V \int_{E_{\min}}^{E_{\max}} \frac{\alpha(E)}{\ln(b/a)E^2} dE \right\} \quad (4.25)$$

If we use the linear approximation that $\alpha(E) \approx \beta E$, where β is an empirical constant, then we can integrate eqn 4.25. Taking $E_{\min} = I/\lambda e$ and $E_{\max} = V/[\ln(b/a)a]$, we can show that

$$G = \exp \left\{ \frac{\beta V}{\ln(b/a)} \ln \left(\frac{V \lambda e}{a I \ln(b/a)} \right) \right\} \quad (4.26)$$

This allows us to understand the rapid and approximately exponential rise of the gas gain with applied voltage that we saw in Fig. 4.12 for the proportional regime.

Proportional wire chambers

Figure 4.13 shows the geometry of a cylindrical proportional chamber. In a typical arrangement, there is a thin anode wire at a high-voltage (HV) potential of a few kilovolts on the axis and the cylindrical cathode is at ground potential. The wire has a radius of 10–20 μm . Assuming that the length of the wire is much greater than its diameter (a very good approximation), we can easily calculate the magnitude of the electric field from Gauss's law. Taking a cylindrical surface around the wire, we can show that the magnitude of the electric field is given by

$$|\mathbf{E}| = \frac{V}{\ln(b/a)r} \quad (4.27)$$

(see Fig. 4.14), where V is the potential difference between the anode and cathode, a and b are the radii of the anode and cathode, respectively, and r is the radial distance from the centre of the anode wire. The cell is filled with a gas. A common, cheap, and safe choice for the gas is a 9 : 1 argon and CO_2 mixture: the noble gas has the advantage of chemical inertness, so the electrons liberated by ionization will be able to travel without being absorbed (the role of the CO_2 is explained below).

A charged particle crossing the cell ionizes the gas, creating about 40–60 electron–ion pairs per centimetre. This number of electron–ion pairs is then increased by a factor of 2–3 because some electrons have enough energy to ionize the gas further. Electrons drift towards the anode and the much slower (massive) ions drift towards the cathode in a diffusion-like process. Very close to the anode, a few times the anode wire diameter, the electric field is high enough (the anode wire is thin) to accelerate the drifting electrons to energies allowing further gas ionization, and this leads to an avalanche process (see the discussion of gas gain earlier in this section).²⁸ There is also recombination of electrons and ions with emission of UV photons. These photons, if not absorbed, could eject electrons from the cathode, leading to a continuous electric discharge. The role of the CO_2 (or another gas with molecules with many degrees of freedom) is to absorb the UV photons and transform their energies to molecular vibration or rotation, which then decay via emission of longer-wavelength photons. These longer-wavelength photons have too low an energy to eject electrons from the cathode. In this typical arrangement, the cell operates in the proportional regime.

Many different types of ‘wire’ chambers have been developed. They do not necessarily even have to contain wires, but they all rely on a large electric field to create an avalanche and they detect the induced currents caused by the drifting electrons and positive ions. These are described in the references in Further Reading.

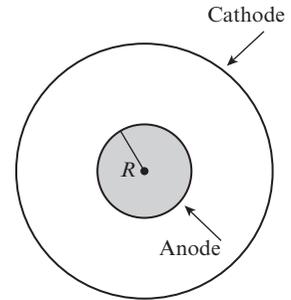


Fig. 4.13 A fundamental cell of a wire chamber. Not to scale.

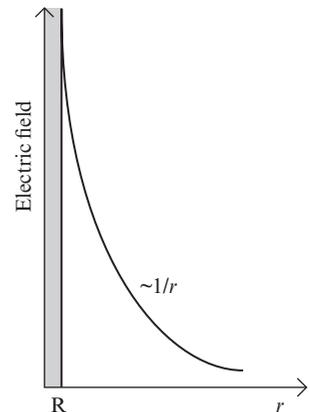


Fig. 4.14 Electric field inside the fundamental cell of a wire chamber.

²⁸Electrons and ions are accelerated by the electric field, but also undergo many collisions with gas atoms and thereby acquire a uniform ‘drift velocity’ superimposed on their random motion as in a conductor. For our purposes, we can ignore the random motion and just consider the drift velocity. However, the random motion contributes to diffusion and is one of the factors limiting the resolution of wire chambers.

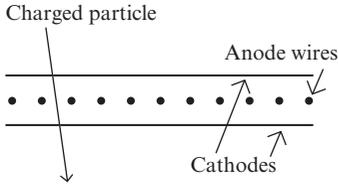


Fig. 4.15 A multi-wire proportional chamber (MWPC).

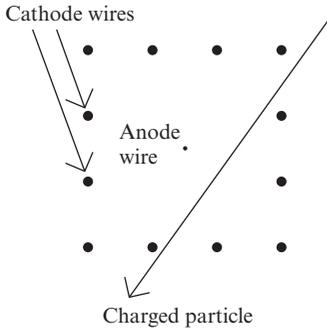


Fig. 4.16 A drift chamber and its fundamental cells.

We will consider two common types of wire chambers:

- multi-wire proportional chambers (MWPCs);
- proportional drift chambers (drift chambers).

A typical example of an MWPC, as sketched in Fig. 4.15, consists of a plane of anode wires between two planes of cathodes (sometimes cathode wires). Such chambers are often used in fixed-target experiments where charged particles are crossing chambers close to perpendicular to their anode planes. If the spacing between anode wires is d , and a simple binary readout is used (i.e. a wire records either a hit or a no-hit), then the resolution of reconstructed points on the charged-particle trajectory is $d/\sqrt{12}$ (see Exercise 4.12). The separation between anode wires cannot be too small, because of the large electrostatic forces on the wires. The wires are held under tension to prevent neighbouring wires touching, but this imposes a minimum separation of a few millimetres.

Drift chambers have better spatial resolution, down to about $50\ \mu\text{m}$. A drift chamber in the barrel of a collider (head-on collisions) detector has a cylindrical structure. The anode wires are parallel to the chamber axis (parallel to the beam direction) and each wire is surrounded by cathode wires, creating a fundamental cell as sketched in Fig. 4.16. Such a cell might be several centimetres across, so the anode (or sense) wires are far apart from each other in comparison with an MWPC arrangement. This arrangement provides position measurements in the plane perpendicular to the beam axis. The trick is to measure the electrons' drift time. Using a signal from a fast independent detector like a scintillator, measuring precisely when particles in colliding beams interacted producing charged particles crossing the drift chamber, one can measure the time between the primary ionization and the leading edge of a signal from an anode wire.

Measurement of position along the beam direction can be done by different techniques. One method is to use anode wires at a small angle to the beam direction, which allows 'stereo' reconstruction of the distance along the beam axis. The geometry of the fundamental cell as well as the gas composition need to be chosen carefully, so the drift velocity²⁹ of electrons is as uniform as possible across the cell, allowing for precise measurement of the location where the primary ionization took place, calculating it from the drift time and the drift velocity.³⁰ In older experiments, the time of the signal was measured relative to an independent signal from a fast detector like a scintillator. At the LHC, an external timing detector is not necessary, because the LHC machine clock running at 40.008 MHz can be used.

Signals and readout for wire chambers

In this section, we will calculate the induced current in a cylindrical wire chamber. The electrons drifting towards the anode will create an avalanche very close to the anode wire. To a first approximation, we can

²⁹Typically a few cm per μs .

³⁰It is often a little more complicated, because, in the presence of a magnetic field, electrons do not drift along lines of the electric field in the drift chamber but at an angle to them, known as the Lorentz angle.

neglect the induced signal from the flow of electrons because they travel such a short distance. We can then calculate the induced current as the positive ions drift away from the wire to the cathode.

We can easily calculate the current induced by the motion of a single ion in the simple cylindrical wire chamber using eqn 4.21. As this is a two-electrode geometry, we can read off the weighting field from the actual electric field by setting the voltage across the chamber to be 1 V and therefore, from eqn 4.27,

$$|\mathbf{E}_W| = \frac{1}{\ln(b/a)r} \quad (4.28)$$

The drift velocity of the ion, \mathbf{v}_d , is related to the electric field \mathbf{E} by $\mathbf{v}_d = \mu\mathbf{E}$, where μ is the ion mobility. We will assume that the mobility is constant. If the number of electron-ion pairs created by the avalanche from a single primary electron is N_{tot} , then the induced current (eqn 4.21) is³¹

$$I = -N_{\text{tot}}e \frac{v_d}{\ln(b/a)r} \quad (4.29)$$

³¹The signal from a single electron is too small to measure. However, with the large gas amplification, the signal can be measured by a suitable low-noise amplifier.

Substituting for the electric field for this geometry, we get the ion speed as

$$v_d = \frac{dr}{dt} = \frac{\mu V_0}{\ln(b/a)r} \quad (4.30)$$

multiplying both sides by r , we can integrate eqn 4.30 and solve for r :

$$r = a \left(1 + \frac{t}{t_0}\right)^{1/2} \quad (4.31)$$

where $t_0 = a^2 \ln(b/a)/(2\mu V_0)$. Substituting from eqn 4.30 into eqn 4.29, we get

$$I(t) = -N_{\text{tot}}e \frac{1}{\ln(b/a)r} \frac{\mu V_0}{\ln(b/a)r} \quad (4.32)$$

Substituting for r from eqn 4.31 into eqn 4.32, we can calculate the induced current as a function of time:

$$I(t) = \frac{-N_{\text{tot}}e}{2\ln(b/a)} \frac{1}{t + t_0} \quad (4.33)$$

This current flows up to the time (t_{max}) when the positive ions reach the anode. We calculate t_{max} from eqn 4.31 by setting $r(t_{\text{max}}) = b$:

$$t_{\text{max}} = (b^2 - a^2) \frac{\ln(b/a)}{2\mu V_0} \quad (4.34)$$

Calculating t_0 and t_{max} for typical conditions (see Exercise 4.11), we find $t_0 \sim 10$ ns and $t_{\text{max}} \sim 100$ μ s. This pulse shape is illustrated in Fig. 4.17.

³²These currents could otherwise flow from the high-voltage power supply to the ground, owing to occasional electric discharges (sparks), and might melt the wires (which are thin).

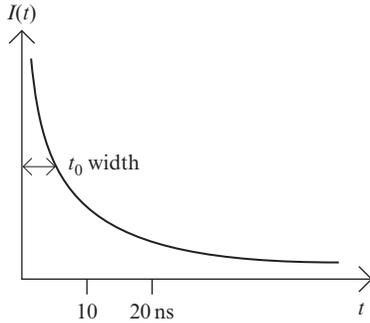


Fig. 4.17 Typical pulse shape from a cylindrical wire chamber.

³³The leakage current represents a noise source that if too large will swamp the signal. In addition, the leakage current will lead to local heating of the silicon, and it is difficult to remove this heat without adding excess material.

³⁴Thermal generation of electron-hole pairs will always occur, but the resulting leakage current is usually acceptable—if it is not, it can be reduced by cooling the silicon.

Equation 4.33 shows that the current pulse has a fast peak and then a very slow ‘tail’. For a high-rate application such as a collider detector, we need fast pulses. We can produce a fast pulse by suitable ‘pulse shaping’; this is done by filtering in frequency space to remove the low-frequency signals. A typical electronic readout circuit is sketched in Fig. 4.18. R_1 is very large ($\sim M\Omega$), thus protecting the anode wires from large currents.³² R_2C_2 and R_2C_1 are time constants, small enough to allow a fast current flow through R_2 , the input resistance of a preamplifier connected to the anode wire, isolated from the high voltage by the C_2 capacitor.

4.6.2 Silicon detectors

Silicon strip detectors as well as silicon pixel detectors are playing an increasingly important role in tracking. The operation of silicon detectors is based on the fact that silicon is a semiconductor with a bandgap of 1.1 eV. A high-energy charged particle traversing silicon will interact with the silicon to create electron-hole pairs. However, most of the energy goes into phonons, so the average energy lost per electron-hole pair created is significantly larger, about 3.6 eV. This results in about 80 electron-hole pairs per micrometre for a minimum-ionizing particle. If no external field were applied, the electron-hole pairs would move apart slowly owing to diffusion; however, this process is too slow for most applications in particle detectors. Therefore, an electric field is applied to separate the electrons and holes. This motion of electrons and holes causes an induced current to flow in the external circuit as discussed in Section 4.4

Even in high-purity, high-resistivity silicon, the presence of a strong electric field would result in an unacceptably large leakage current, i.e. current flowing even without the presence of the charged particle.³³ This problem is solved by making a *pn* junction, which forms a diode junction. When a reverse bias is applied to the diode, the free electrons are removed from the *n*-doped region, creating a ‘depletion’ region, in which there is a very low density of free carriers, thus allowing a large electric field to be applied, without paying the price of the unwanted large leakage current.³⁴ How thick does the silicon have to be to create a big enough signal? There is actually no correct answer to this question, because it depends on the amplifier, but a typical choice is 300 μm , which

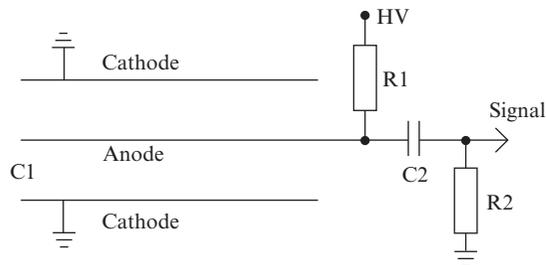


Fig. 4.18 Fundamental cell readout circuit.

results in a signal of about 25 000 electron–hole pairs for a minimum-ionizing particle. The next question to consider is how large an electric field is needed to fully deplete the silicon. We can answer this question starting from Poisson’s equation for the potential V in terms of the charge density ρ and the dielectric constant ϵ :

$$\nabla^2 V = -\rho/\epsilon \quad (4.35)$$

If we are assuming an effectively one-dimensional diode, and N is the net volume number density of charges, we can use eqn 4.35 to calculate the potential as a function of the distance x :

$$\frac{d^2 V}{dx^2} + \frac{Ne}{\epsilon} = 0 \quad (4.36)$$

where e is the electron charge and ϵ is the permittivity of silicon. We will consider a detector with p strips in n bulk silicon. In this case, the p region is much more heavily doped than the n region, so we only need to consider the n -doped region. On applying the reverse bias, we remove all the free electrons from the n -doped region, which leaves behind a fixed space-charge density. Integrating eqn 4.36 gives³⁵

$$\frac{dV}{dx} = -\frac{N_d e}{\epsilon}(x - x_n) \quad (4.37)$$

³⁵ dV/dx is just equal to minus the electric field.

where N_d is the donor (electron) density and x_n is the limit of the depletion region. Integrating eqn 4.37 gives

$$V = -\frac{N_d e}{\epsilon} \left(\frac{x^2}{2} - x x_n \right) \quad (4.38)$$

Finally, the total voltage applied across the depletion region is found by setting $x = x_n$:

$$V_{\text{bias}} = \frac{N_d e x_n^2}{\epsilon} \quad (4.39)$$

Equation 4.39 shows why we need high-purity silicon to make good detectors—because impurities contribute to N_d and hence cause an increase in the bias voltage required for full depletion.³⁶ With typical detector-grade silicon, a 300 μm -thick silicon detector requires a bias voltage of about 50 V (see Exercise 4.13) We can calculate the drift velocities for electrons and holes:

$$v_{\text{drift}} = \mu E \quad (4.40)$$

where μ is the mobility and E is the electric field. We can use the measured mobilities to calculate the maximum drift times for electrons and holes (see Exercise 4.13). Detectors are typically operated at higher bias voltages to speed up the signal collection. Great care is needed in the design of silicon detectors, because too large electric fields can lead

³⁶Too high a bias voltage will result in electrical breakdown in the cables or the silicon detector itself.

to electrons gaining enough energy to cause secondary ionization, which leads to an avalanche effect and hence results in electrical breakdown. This will start in the region of highest electric field; any very small-scale non-uniformities in the electrode structure can cause enhanced electric fields and hence lead to electrical breakdown, even at relatively low bias voltages.

³⁷In normal operation, there is no equivalent of gas gain in silicon detectors, although devices like avalanche photodiodes can be operated at sufficient voltage for amplification to occur.

³⁸Occurring every 25 ns under nominal LHC operating conditions.

³⁹Electrical power consumption must be minimized: if more heat is generated, the cooling system must be larger, degrading the tracker resolution and creating unwanted secondary interactions in the tracker.

⁴⁰Gy(Si) is the SI unit of dose, corresponding to 1 J of energy deposited per kg of Si: a dose of 100 kGy(Si) corresponds to $\sim 10^9$ lung X-rays.

To determine whether this small signal³⁷ can be detected, it is essential to consider all sources of electronic noise. This is discussed in detail in Spieler's book in Further Reading, from which we see that we need to have low-capacitance detectors. For high-rate applications, such as the silicon trackers at the LHC, we need to minimize 'pile-up' backgrounds from hits in previous bunch crossings³⁸ generating spurious hits in the triggered bunch crossing. This implies that the 'shaping time' of the electronics should be not more than $O(25 \text{ ns})$. The challenge is to design low-noise amplifiers that are sufficiently fast and consume low power.³⁹

Radiation damage

One of the difficulties with the application of silicon detectors in particle physics, particularly at the LHC, is radiation damage. At a radius of 30 cm from the beam line, the expected ionizing dose over the detector lifetime is 100 kGy(Si).⁴⁰ High-energy particles can displace silicon atoms from their lattice sites, creating complex defects that result in states between the valence and conduction bands (called mid-bandgap states). This makes it much easier for thermal generation to promote an electron from the valence to the conduction band. This greatly increases the leakage current. The leakage current is strongly dependent on the temperature T :

$$I_{\text{leak}}(T) = AT^2 \exp\left(-\frac{E_g}{2k_B T}\right) \quad (4.41)$$

where k_B is Boltzmann's constant and E_g is the bandgap, which for silicon is 1.1 eV.⁴¹ Therefore, the leakage current can be very efficiently suppressed by cooling the silicon. These mid-bandgap states act like extra acceptors and thus change the effective dopant concentration.⁴² From eqn 4.39, we can see that an increase in the effective dopant concentration will result in detectors requiring higher bias voltages to be fully depleted. The electrical breakdown of detectors at very high voltages therefore sets the scale for the maximum radiation doses that can be tolerated. In addition, some of the extra states can cause 'charge trapping', which looks like a signal loss.

Silicon systems

The spatial resolution of a silicon detector is largely determined by the segmentation of the silicon into individual detector channels. If the width of a detector segment is x , and if a particle only causes a hit in a single

⁴¹A useful rule of thumb is that the leakage current doubles for every 7 K increase in temperature.

⁴²With sufficient damage, this can cause n -type silicon to change to p -type silicon, a process called type inversion. However, detectors can operate satisfactorily after type inversion.

channel, then the spatial resolution in this direction is $x/\sqrt{12}$. Improved resolution can be achieved by using signals in neighbouring channels; the amount of charge sharing with neighbouring channels gives extra information on the location of the ‘hit’.

There are generally two classes of silicon detector systems: strips and pixels.

Strip systems

A very simplified schematic cross-section of part of a generic silicon strip detector is shown in Fig. 4.19. A positive high voltage is applied to the ‘backside’ via the Al contact, which depletes the n -bulk silicon. Electron–hole pairs created by ionizing particles drift in the electric field and the current induces signals on the readout electrodes. The signal electrodes are AC-coupled to the Al strips (using the SiO_2 as an insulator), which are then connected to the preamplifiers in the readout ASIC (application-specific integrated circuit). The noise increases with the detector capacitance (see Spieler in Further Reading for an explanation); therefore, for high-rate applications such as the LHC, we must minimize any stray capacitance between the detector and the amplifier. The connection is made with ‘wire bonds’, typically a few millimetres long and $25\ \mu\text{m}$ -thick aluminium wire. These thin bond wires can be ultrasonically bonded to pads on the detector and on the readout ASIC. This allows a very short connection between the detector and the amplifier, which introduces much less capacitance than a longer wire cable. We also need to create a DC return path for the current and this requires a large-value resistor, so that the fast signal flows across the capacitor. This is achieved with polysilicon resistors inside the silicon detector itself.

In a strip detector, the silicon wafer is divided into long narrow strips, with typical strip widths in the range of $50\text{--}100\ \mu\text{m}$ (the largest wafers used are 6 inches in diameter). This is done to obtain good precision in the bending plane of the magnetic field. Modest resolution ($\sim 1\ \text{mm}$) in the orthogonal direction is achieved by having half the sensors with a small stereo angle. This has the disadvantage that it creates ambiguities if more than one particle hits a sensor.⁴³

The amplifiers are in custom-designed ASICs. As the time taken for the first-level trigger (L1) (see Section 4.10) is of the order of microseconds, which is much longer than the $25\ \text{ns}$ between bunch crossings, the data must be kept on-detector until the trigger decision is made. This is achieved with ‘pipeline’ memory in which the data from each strip for each bunch crossing are stored in different memory elements (see Fig. 4.20). If the L1 rejects the event, the corresponding data can be overwritten. If the event is triggered at L1, the corresponding data are read out via optical links.

A schematic view of an ATLAS Semi-Conductor Tracker (SCT) module is shown in Fig. 4.21. The module consists of two pairs of silicon wafers glued together to make a double-sided module. The ASICs are mounted on flexible copper–Kapton circuits. The beryllia (BeO) ‘ear’

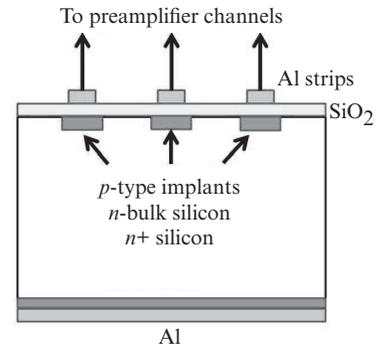


Fig. 4.19 Schematic cross-section through a silicon micro-strip detector with p implants in an n -bulk silicon.

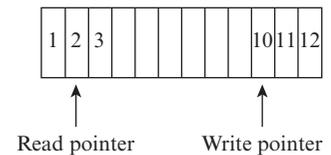


Fig. 4.20 Principle of a pipelined memory. At each clock cycle, data are written into the cell defined by the write pointer. This pointer is advanced by one cell every clock cycle; after it gets to the last cell, it cycles back to the first. The read pointer follows a fixed number of clock cycles behind the write pointer. The time delay between the write and read pointers defines the time available for making a trigger decision. If the decision is positive, the data are read out from the corresponding cell; if not, then new data can be read into this cell. When the pointers advance beyond the last cell (12 in this unrealistic example), they cycle back to cell 1.

⁴³In the ATLAS case, a discriminator is used to determine if hits are above threshold, so the output data are digital. In the CMS tracker, the signal amplitude is transmitted off-detector via analog optical links.

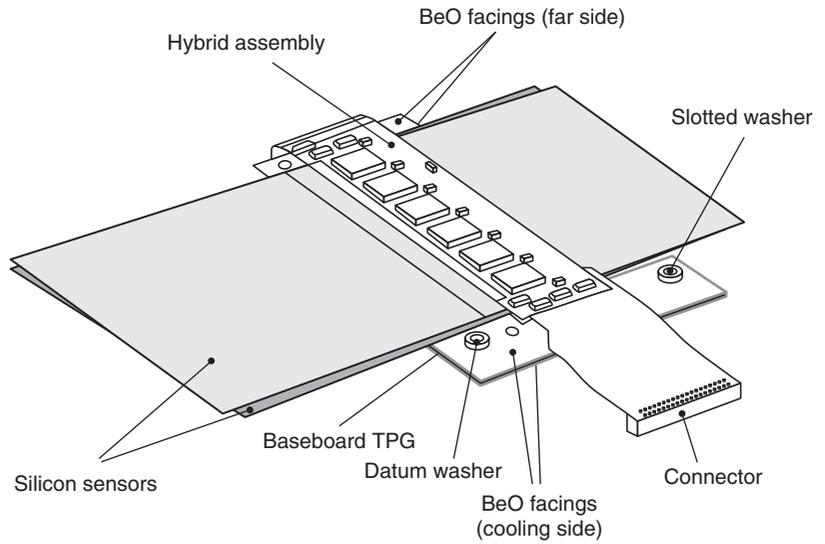


Fig. 4.21 Schematic view of an SCT module [6].

at the side allows the module to make good thermal contact with the cooling tube. The coolant used is perfluoropropane (C_3F_8), since this provides very efficient two-phase cooling; i.e. the heat from the ASICs and the silicon detectors is used to evaporate liquid C_3F_8 . These are very large systems, with 60 m^2 of silicon detectors for ATLAS and 200 m^2 for CMS. The modules have to be held rigidly in place to benefit from the high intrinsic spatial resolution, but the material must be minimized because any material causes multiple scattering of all charged particles and results in electrons and photons starting electromagnetic showers before the calorimeter. Therefore, each module is mounted on carbon-fibre support structures since these provide the best ratio of stiffness to weight.

Pixel systems

In silicon pixel detectors, the silicon is divided into much smaller areas; for example, in the ATLAS pixel detector, the dimensions of individual pixels are $50\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$. The smaller dimension is in the bending plane of the magnetic field to optimize the momentum resolution. The first advantage of pixel over strip detectors is that they provide unambiguous high-precision space points. In addition, the ‘occupancies’ (i.e. the fractions of detector elements that are hit in given events) are much lower for pixel detectors than for strips. This is vital for pattern recognition at the LHC, which has to reconstruct tracks in the presence of ‘pile-up’ background from about 25 collisions in the same bunch crossing. The small area of the pixels means that the detector capacitance is very low, which allows very low noise to be achieved (see Spieler in Further Reading). However, this requires minimization of stray capacitance between the silicon pixel and the amplifier in the ASIC. One of

the main difficulties with pixel systems is how to make the electrical connection from each silicon pixel to a unique channel of the readout ASIC without introducing any significant capacitance. This is achieved by ‘bump bonding’.⁴⁴ The much larger number of channels in pixel systems than in strips requires more sophisticated data processing in the ASICs.⁴⁵ Other system aspects for pixels are similar to those for strips.

Pixel systems offer many performance advantages over strips, but as the electronics covers essentially the full sensitive area, a layer of pixel detector will have more material than an equivalent layer of strips. In addition, pixel detectors are significantly more expensive than strip detectors of the same dimensions. Therefore, LHC detector systems are a compromise, with pixels being used close to the beam pipe and strips being used further away.

4.6.3 Tracker performance

Consider the track of a charged particle with momentum p (measured in GeV) perpendicular to a magnetic field \mathbf{B} . The radius of curvature R is related to the momentum by $p = 0.3BqR$. We assume that the track is measured over a length l (see Fig. 4.22). From the geometry, we can relate R to the ‘sagitta’ s and l by Pythagoras’ theorem: $R^2 = (R - s)^2 + (l/2)^2$. For high-momentum tracks, we can neglect the s^2 term and find $1/R = 8s/l^2$. Therefore, the error in $1/R$ is given by $\sigma(1/R) = 8\delta s/l^2$. To make approximate estimates of the momentum resolution, we will assume that the track is measured very precisely at the start and end of the trajectory but with an error given by δs at the midpoint. In this approximation,

$$\sigma(1/p) = \frac{8\delta s}{0.3Bql^2} \quad (4.42)$$

Although eqn 4.42 is a rough approximation, some general features are valid:

- (1) **B field:** The resolution improves with the value of \mathbf{B} , so we wish to use the largest value possible. Using superconducting magnets, fields up to 4 T have been achieved.
- (2) **Length:** The resolution improves as l^2 ; however, for a tracker in a collider detector, the value of l is set by the inner radius of the electromagnetic calorimeter. Increasing l too far is therefore impractical for cost reasons—for the ATLAS inner detector, the track length is about 1 m. For muon spectrometers, the constraints are weaker and large values of l can be used (e.g. in the ATLAS muon spectrometer, $l \sim 5$ m).
- (3) **Scaling:** The resolution is constant in $1/p$, which implies that the momentum resolution degrades with increasing momentum.

If \mathbf{B} and l are fixed and we wish to measure momenta up to some value p_{\max} , we can use eqn 4.42 to estimate the required spatial resolution

⁴⁴In this process, indium solder is deposited on metallized pads on the pixel and heated in a reflow process to form hemispherical solder balls; the detector is then flipped and positioned very precisely over a flexible circuit with the readout ASICs already mounted. A further reflow of the solder results in an electrical connection between the pixels and the amplifiers in the readout ASICs. It is difficult to achieve a high yield and this process is very expensive.

⁴⁵The area required in an ASIC to implement a pipeline for each pixel would be unacceptably large. Therefore, another approach is used for the pipeline that benefits from the very low occupancies in the pixels. A data-driven pipeline is used so that when a pixel is above threshold for a given bunch crossing, a ‘time stamp’ for that pixel address is written in memory. When a first-level trigger is received, the data for all the pixels with the correct time stamp are read out.

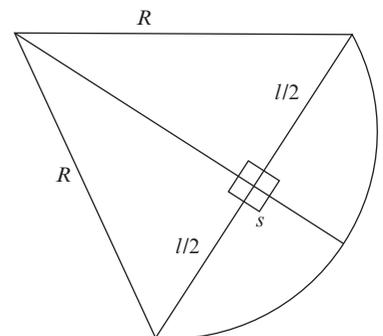


Fig. 4.22 Definition of the track sagitta s .

(see Exercise 4.10). So far, we have only considered the contribution of the precision of the measurement points. However, in a real detector, we have material, so the charged particles undergo multiple scattering (eqn 4.6). As the scattering angle is inversely proportional to the momentum, this causes a contribution to the error in p that is constant. Therefore, the momentum resolution of a real tracker can be parameterized by adding the effects of measurement precision and multiple scattering in quadrature:

$$\sigma(1/p) = C \oplus D/p \quad (4.43)$$

where D is the term due to multiple scattering and C is the term due to measurement error resolution.⁴⁶

Another important measure of the performance of a tracking detector at a collider is how precisely the tracks can be extrapolated back to the primary vertex. Particles originating from decays of b or c quarks or τ leptons will travel for the order of 1 ps before decaying, and hence if one extrapolates the tracks back, they will miss the primary vertex. In the plane transverse to the beam direction, this distance is called the impact parameter (see Fig. 4.23). The resolution in impact parameter depends on the intrinsic resolution of the tracker and multiple scattering. Therefore, one requires a very high-precision measurement as close to the beam line as possible, and this is performed with silicon detectors (either strip or pixels). To minimize multiple scattering, one needs to have a very thin (in radiation lengths) beam pipe, and the best choice is beryllium (although beryllium is very difficult to machine and hence expensive).

⁴⁶The optimization of the overall resolution is an interesting trade-off, because adding more measurements will decrease D but will add more material and therefore increase C .

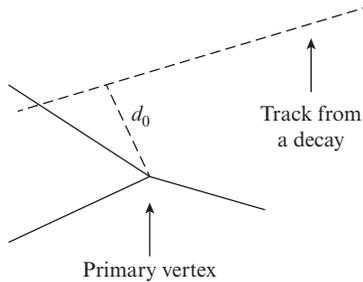


Fig. 4.23 Schematic view of tracks in the transverse plane showing tracks from the primary vertex and the definition of the impact parameter d_0 from the one track resulting from a decay.

⁴⁷We cannot afford to instrument the entire calorimeter with the fine segmentation required to measure the electromagnetic showers.

4.7 Detectors for particle jets

The energies of particles and ‘jets of particles’ are measured in detector systems called ‘calorimeters’. Ideally, all particles with the exception of muons and neutrinos (or still to be discovered neutrino-like weakly interacting particles) should deposit all their energies in the calorimeter. As electromagnetic showers occupy much smaller volumes than hadronic showers (see Section 4.3), we require much finer segmentation for the front of the calorimeter than the back.⁴⁷ Therefore, the design of calorimeter systems is usually split into ‘electromagnetic calorimeters’ and ‘hadronic calorimeters’.

4.7.1 Electromagnetic calorimeter

The depth of the electromagnetic calorimeter is chosen such that nearly all the energy of electromagnetic showers from electrons and photons of the interesting energy range is contained in this part of the calorimeter. This can be determined from Monte Carlo simulations such as those illustrated in Fig. 4.6. At LHC energies, we need to measure electrons

and photons with energies of several hundred GeV; therefore, the electromagnetic calorimeter needs to be about $25X_0$ deep. Finer longitudinal sampling will also help separate showers induced by electrons from those induced by hadrons. The lateral shower size is set by the Molière radius (see Section 4.3), which for lead is $R_M = 1.8$ cm. The scale for the lateral size of hadronic showers is set by the hadronic interaction length λ_I and is typically an order of magnitude larger. We can therefore achieve further separation between showers induced by electrons and hadrons with fine lateral and longitudinal segmentation. There are two different types of electromagnetic calorimeters:

- ‘sandwich’ calorimeters with alternating layers of ‘active’ and ‘passive’ material;
- homogeneous calorimeters in which one material fulfils the function of absorber as well as actively detecting the presence of the shower.

There are many trade-offs between these approaches. In a sandwich calorimeter, most of the energy is deposited in the passive layers and there are significant fluctuations in the fraction of the energy deposited in the active layers. This usually limits the resolution of sandwich calorimeters and the best resolution can be achieved with homogeneous calorimeters, for which this effect does not arise. However, the average density of crystals used in homogeneous calorimeters tends to be lower than that in sandwich calorimeters, which therefore increases the depth of the electromagnetic calorimeter. This results in larger volumes for the hadronic calorimeter and the muon system, and will thus increase the cost.

4.7.2 Homogeneous calorimeters

Homogeneous calorimeters are usually based on scintillating crystals (Section 4.4.2). The CMS electromagnetic calorimeter (ECAL) is an example of this technique. At the LHC, the scintillation must be fast because of the short time between bunch crossings (25 ns). The crystals must have very good radiation tolerance in order to survive many years of LHC operation. Finally, the crystals must have a very high density in order to keep the dimensions small enough. The CMS electromagnetic calorimeter is based on lead tungstate (PbWO_4) crystals with a density of 8.28 g cm^{-3} and a radiation length of 0.89 cm. About 80% of the scintillation light is emitted in less than 25 ns [62]. One challenge with this system is that the transparency of the crystals decreases with radiation, and therefore sophisticated monitoring techniques are required to compensate for these effects. In addition, the light output is very sensitive to temperature, so the temperature needs to be maintained at a constant value. Because photomultipliers cannot be used in very strong magnetic fields, the scintillation light is read out by avalanche photodiodes (APDs).⁴⁸ A photograph of one such crystal with the APD readout is shown in Fig. 4.24.

⁴⁸In the end-cap calorimeter, the radiation levels are too large for the use of APDs, and vacuum phototriodes are used instead.



Fig. 4.24 Photograph of a PbWO_4 crystal and readout for the CMS electromagnetic calorimeter [62].

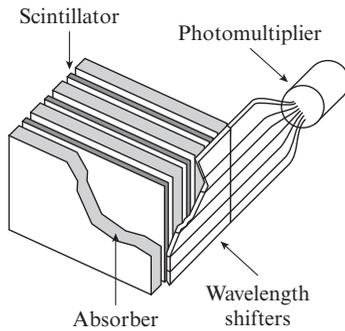


Fig. 4.25 Schematic view of one cell of a sandwich scintillator calorimeter with wavelength-shifting plates to guide the light to the photomultiplier at the back.

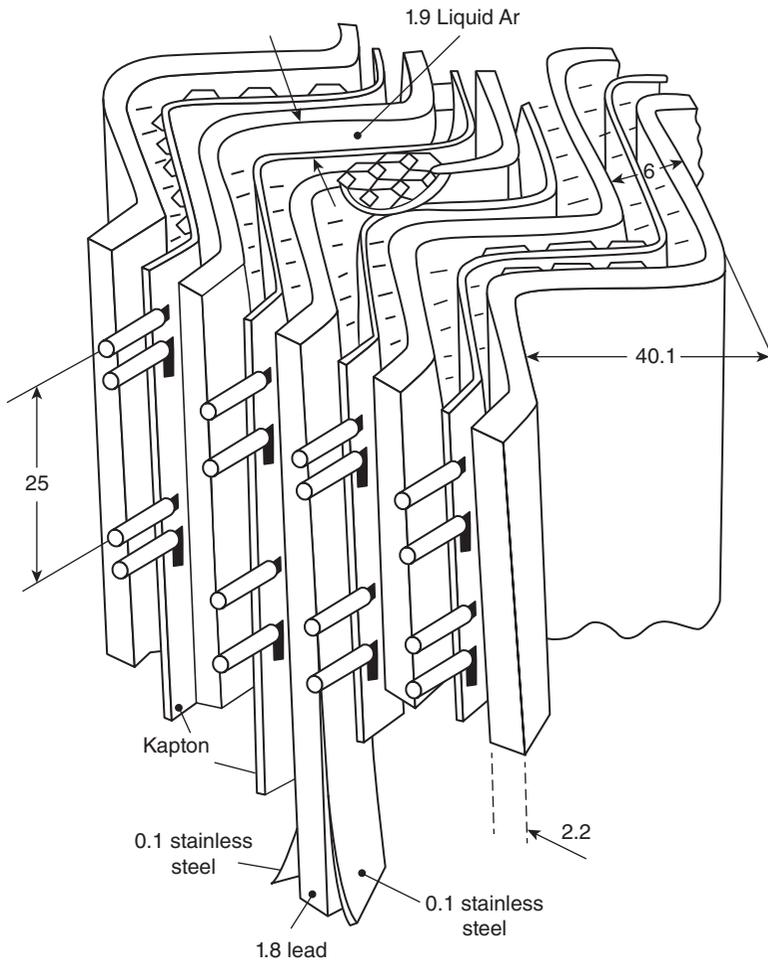
4.7.3 Sandwich calorimeters

In a sandwich calorimeter, there are alternating layers of active and passive material. The passive material should have high Z (to enable a relatively compact design)—lead is a common choice. The total energy detected in the active layers is only a fraction of the total energy deposited. This fraction can be measured in prototypes or small parts of the calorimeter in dedicated test beams in which the energy of the incident electrons is fixed, although at the LHC the rate of Z production is so high that ‘in situ’ calibration can be performed. Before the LHC, the most common design of electromagnetic sandwich calorimeter used plastic scintillators for the active layers. The scintillation light (see Section 4.4.2) needs to be guided to the photomultipliers at the back of the calorimeter. This is done using ‘wavelength-shifting’ plates (see Fig. 4.25). These contain fluors to shift the wavelength to longer wavelengths (typically in the green), for which the plastic is more transparent. There are several limitations to this technique:

- **Cracks:** Each tower requires a support structure to hold it in place which introduces dead zones between cells (called ‘cracks’).
- **Non-uniformity:** The absorption length for the scintillation light is typically the same magnitude as the lateral dimensions of the cell; therefore, the response will depend on the impact point of the electron.
- **Radiation damage:** The scintillator will suffer significant radiation damage, and very good calibration schemes are essential to track this. The most common method used is to move a radioactive source such as ^{60}Co over the calorimeter.

A newer approach to scintillator sandwich calorimeters uses wavelength-shifting fibres embedded in the scintillator to transport the light to the photodetectors. This avoids the need for bulky waveguides, which add to the ‘cracks’ between calorimeter cells.

To overcome these limitations, a novel type of electromagnetic calorimeter has been developed for ATLAS, based on a new geometry for lead absorbers and liquid-argon ionization chambers. The signals are generated by electrons created by ionization, drifting in a large electric field and generating an induced current at the electrodes (see Section 4.4).⁴⁹ The fundamental problem with this technique for use at the LHC is that typical drift times for the electrons are ~ 400 ns, which is much longer than the time between bunch crossings of 25 ns. The solution is based on very fast ‘bipolar’ pulse-shaping electronics, in which most of the signal is not detected but a suitably fast pulse is generated. This is illustrated in Fig. 4.26. As most of the signal is not utilized, it is essential to lower the noise in order to maintain the signal-to-noise ratio. This is achieved by lowering the capacitance and inductance of the electrodes using a novel ‘accordion’ geometry as shown in Fig. 4.27. An important advantage of this technique is that liquid argon is inherently radiation-hard.



⁴⁹This is very similar to a wire chamber operating in the ‘ionization chamber’ region, in which there is no gas gain. However, as liquids are much denser than gases, a high-energy charged particle can create sufficient ionization for the signal to be detectable.

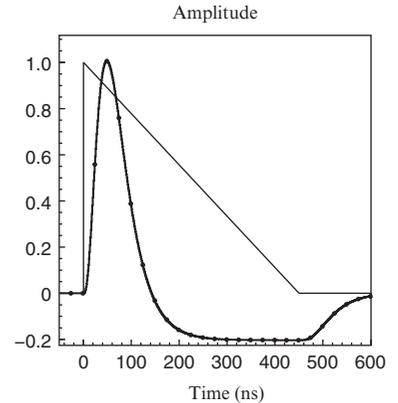


Fig. 4.26 Signal pulse shape in the ATLAS liquid-argon calorimeter [20]. The triangular shape is the current pulse created by the electron drift. The curve shows the pulse shape after shaping with a bipolar pulse shaper.

Fig. 4.27 Sketch of a small section of a prototype for the ATLAS electromagnetic calorimeter, illustrating the ‘accordion’ structure [38] (all dimensions are in millimetres). In this geometry, the signals are transported to the electronics on flat copper/Kapton tapes, which have lower capacitance and inductance per unit length than the cables that would be required if the electrodes were orthogonal to the direction of incidence of particles. The absorber plates are made from lead lined with stainless steel. The liquid argon is contained between the absorber plates, and the copper/Kapton electrodes are attached to these plates.

	ATLAS	CMS
a (% GeV ^{1/2})	10.0	2.8
b (GeV)	0.4	0.12
c (%)	—	0.3

Table 4.2 Electromagnetic calorimeter resolution for prototype calorimeters measured in test beams.

It is relatively easy to divide the readout cells to the desired lateral and longitudinal granularity. Another critical advantage is that the structure is self-supporting, so there is no need for passive material between cells, thus avoiding the cracks inherent in calorimeters based on plastic scintillators for the active layers.

4.7.4 Resolution

The energy resolution of a typical electromagnetic calorimeter can be parameterized as

$$\frac{\Delta E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (4.44)$$

where a , b , and c are constants and the different terms are added in quadrature. The constant a represents the ‘stochastic term’, b represents the contribution from electronic noise, and c is a constant term. In a calorimeter using a scintillator, if at a given energy the mean number of detected photons is N , there will be Poisson fluctuations giving a contribution to the stochastic term

$$\frac{\Delta E}{E} \sim \frac{\Delta N}{N} \sim \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}} \Rightarrow \frac{\Delta E}{E} = \frac{a}{\sqrt{E}}$$

However, in a sandwich calorimeter, this effect is usually negligible compared with the ‘sampling’ fluctuations, i.e. the fraction of energy deposited in the active layers.⁵⁰ The constant b in eqn 4.44 represents the contributions from electronic noise and should be negligible at high energies. The constant term c represents the effects of residual non-uniformities in response across the cell and over all cells, as well as variations in time. With the aid of good calibration procedures, the constant term can be reduced to less than 1%.⁵¹ Measured parameters from test-beam studies of the ATLAS [20] and CMS [62] electromagnetic calorimeters are given in Table 4.2. However during LHC operation, there are other factors that will degrade the resolution, such as radiation damage, uncertainties in the calibration constants, and ‘pile-up’ backgrounds (particles from extra collisions in the same bunch crossing). For the very

⁵⁰There are many interesting trade-offs here. If the scintillator/passive ratio is increased, sampling fluctuations are reduced, but the size and cost of the calorimeter are increased. As discussed in Sections 4.7.1 and 4.7.2, there is no perfect design.

⁵¹At the LHC, the very large sample of $Z \rightarrow e^+e^-$ events provide ample data for in situ calibration of the electromagnetic calorimeters.

important Higgs decay, $H \rightarrow \gamma\gamma$, the precision of the angular measurement also contributes to the mass resolution. These factors favour the higher granular segmentation, the intrinsic stability, and the radiation hardness of a liquid-argon calorimeter compared with a scintillator calorimeter. The result is that the mass resolution for the Higgs decay $H \rightarrow \gamma\gamma$ is comparable for ATLAS and CMS.

4.7.5 Hadronic calorimeter

The hadronic calorimeter surrounds the electromagnetic calorimeter. Ideally, the combined electromagnetic and hadronic calorimeters should contain nearly all the energy from showers from hadrons entering the calorimeters (mostly π^\pm). An indication of the required depth of the calorimeter can be deduced from the curves in Fig. 4.8. The practical depths for hadronic calorimeters are constrained by cost and available space, but a rule of thumb is that at LHC energies a depth of at least about 10 nuclear interaction lengths is required. A homogeneous hadronic calorimeter would be too large and so is not a practical option, and the hadronic calorimeter will be of the ‘sandwich’ type. The resolution for hadronic calorimeters is greatly reduced if the calorimeter is not ‘compensating’, which means that the ratio of the response to electrons to that of hadrons, e/h , is significantly different from unity (see Section 4.3.6). There are several possible approaches to achieving compensation in hadronic calorimeters:

- **Tuning the ratio of absorber to active thickness:** The energy loss for electrons scales as Z^2 , compared with Z for charged hadrons; therefore, in thicker absorbers, the value of e/h can be lowered (however, this has other problems that will be discussed later). Lower- Z cladding can be used to absorb low-energy photons preferentially, which also reduces e/h .
- **Increasing the hadronic response:** Instead of trying to suppress the response to electrons, we can try to enhance the response to hadrons. There are many low-energy neutrons that can be indirectly detected by elastic scattering off nuclei. The optimal nucleus is hydrogen, so detectors containing hydrogen, such as organic scintillators, can be used.
- **Use of depleted uranium:** One suggestion to increase the hadronic response was to have uranium absorber plates and use the energy released by fission after fast-neutron capture.⁵²
- **Software compensation:** In a finely grained calorimeter, calibration procedures can be optimized to try to achieve compensation; this approach is discussed below.
- **Dual readout:** The idea is to read out the shower energy using two different techniques with very different values for e/h . A prototype of such a hadronic calorimeter has been built by the DREAM collaboration and it uses copper tubes, each filled with scintillator

⁵²This was the motivation for the use of uranium in the ZEUS calorimeter, which achieved compensation. However, the two first items were more important than fission.

and quartz fibres. The signal from the scintillator (S) and the Čerenkov (C) radiation in the quartz fibres are measured separately. The values of e/h are very different for the S and C signals, which enables determination of the electromagnetic fraction f_{em} for individual showers. The effect of e/h being different from unity can therefore be corrected, effectively achieving the good hadronic resolution of compensating calorimeters.

Although compensating calorimeters have been built, there are disadvantages in cost and/or resolution for electrons and photons, and the calorimeters for the LHC experiments are not compensating.⁵³ In a highly segmented calorimeter such as that used by ATLAS, the hadronic resolution can be improved by ‘software compensation’; the secondary electromagnetic showers are smaller than hadronic showers, so they lead to higher energy density in the calorimeter cells. Therefore, the electron response can be decreased by de-weighting cells with large energy, thus making the response closer to being compensating and thereby improving the resolution. If the calorimeter cells are calibrated using electrons, the naive estimate of the energy in a hadronic shower would be given by $E = \sum_i E_i$, where E_i is the energy in the i th calorimeter cell. As electromagnetic showers are more compact, the cells with higher local energy density will probably have arisen from electromagnetic showers. A correction factor is applied for hadronic showers. The correction factor decreases for showers with higher local density of energy deposition.⁵⁴ The calibration procedure used to determine the calibration factors aims to reconstruct the true energy on average and to optimize the resolution.

The resolution for hadronic calorimeters can be parameterized by the same form as for electromagnetic calorimeters (eqn 4.44). The stochastic term will be larger because of the relatively coarse sampling, and if the calorimeter is non-compensating, then there will be a large constant term, which will dominate the resolution at high energies. If the calorimeter is not sufficiently deep, the energy lost at the back of the calorimeter will also contribute to the constant term. Any crack regions between cells or non-uniformity of the response over a cell will also add to the constant term. Typical examples of hadronic resolution for compensating and non-compensating calorimeters are given in Table 4.3. The superior

⁵³The ATLAS barrel calorimeter has $e/h \approx 1.4$.

⁵⁴The correction factors also depend on the energy as well as the local energy density.

	ATLAS tile calorimeter	ZEUS
a (% GeV ^{1/2})	52	35
b (GeV)	1.6	—
c (%)	3.0	—

Table 4.3 Energy resolution for prototype hadronic calorimeters measured in test beams.

resolution of the compensating ZEUS calorimeter [51] compared with the non-compensating ATLAS scintillating tile calorimeter [20] is clear, but even so the resolution is far inferior to that achieved by electromagnetic calorimeters. However, the compensation achieved in the ZEUS calorimeter came at the price of degrading the electromagnetic resolution. So, as is usual in detector physics, there is no perfect answer and designs must be optimized to the requirements of a particular experiment.

4.8 Detectors for particle identification

In this section, we review some detector techniques for particle identification. Some particle identification is performed by combining signals in different types of detectors,⁵⁵ but here we restrict ourselves to types of detectors that give standalone particle identification.

⁵⁵For example, a high-momentum track that is matched to an electromagnetic shower in a calorimeter can be identified as an electron.

4.8.1 Particle identification with Čerenkov detection

There are two practical applications of Čerenkov radiation for particle identification:

- **Threshold counter:** If we measure the momentum p of a charged particle, we can determine its speed depending on what particle type it is (and hence what mass it has). For some range of momentum, we can arrange that $v > 1/n$ for one type of particle (e.g. π^\pm) but is below threshold for another (e.g. K^\pm). Therefore, we can separate π^\pm from K^\pm depending on whether a Čerenkov signal was detected.
- **Ring imaging Čerenkov (RICH):** A RICH detector represents a more sophisticated use of Čerenkov radiation in which we measure the direction of Čerenkov photons. This requires optics to focus the photons of a given angle to a particular location on the photon detector. We then associate particular Čerenkov photons with particular charged particles and fit a ring (hence the name of the technique) and measure the Čerenkov angle. If we know the refractive index of the medium, we can then determine the speed of the charged particle. Knowing the momentum p from an independent detector then allows us to estimate the mass of the charged particle and hence identify it as a pion, kaon, etc We will look at an example of a RICH detector in Chapter 10.

4.8.2 Particle identification with transition radiation

We have seen that charged particles crossing a boundary between two dielectric layers can emit X-rays. As the transition radiation increases

⁵⁶This is not a problem for fixed-target experiments; however, it is very problematic for collider detectors, for which the radial space for the tracker is limited by the inner radius of the calorimeter.

⁵⁷The problem is that there are large statistical fluctuations in the magnitude of the energy loss deposited by ionization in a short path length in a gas.

with the Lorentz γ factor, for practical purposes the yield is only significant for high-energy electrons, and this therefore provides a method to separate high-energy electrons from charged hadrons. As the photon yield per dielectric boundary is so low, we need many such boundaries. This sets a lower limit on the required length for a useful transition-radiation detector.⁵⁶ The transition-radiation photons are in the X-ray region. These X-ray photons can be detected in wire chambers with a large fraction of a heavy noble gas like xenon. Xenon has $Z = 54$, which results in a large absorption cross section for X-rays, thus increasing the probability of X-ray absorption in a thin layer of gas. The energy deposited by X-rays is larger than the typical energy deposited by ionization in the gas, so a suitable discriminator level can be set that is sensitive to the X-rays from transition radiation but is rather insensitive to ionization.⁵⁷

4.8.3 Particle identification with ionization

We saw that the rate of energy loss by ionization depends on the speed β of the particle (see eqn 4.5). Therefore, if we can make a suitable precise measurement of the energy loss by ionization and the momentum of a particle, we can achieve some separation between particles with different masses (e.g. pions and kaons). The momentum can be measured by a tracking detector in a magnetic spectrometer, and the amplitude of the signals in the elements of the tracking detector provide a measurement of the energy loss by ionization. The first difficulty with this technique is the presence of very large fluctuations in energy loss by ionization in thin layers, so if a wire chamber is used, a very large number of samples is required to achieve useful particle identification. The second problem is that the rate of energy loss as a function of momentum ‘plateaus’ at high momentum, so this technique is only useful at lower energies.

4.9 Magnetic fields

We need magnetic fields for trackers and muon spectrometers in order to use the measured trajectory to reconstruct the momenta. The magnets are usually based on the same NiTi superconducting technology discussed for accelerators in Section 3.3. The volumes of the magnets are very much larger and, although the magnetic fields are smaller, the energy stored in these fields is very much greater, which leads to new engineering challenges.

4.9.1 Magnetic fields for trackers

The usual choice of field configuration for trackers at colliders is a solenoid (with the axis along the beam line). To minimize the volume and cost, one option is to place the solenoid between the tracker

and the calorimeter. Clearly, too much ‘passive’ material upstream of the calorimeter will degrade the resolution of the electromagnetic calorimeter. Therefore, the fields are generated using superconducting magnets, with field strengths up to 2 T being typical. The CMS magnet has a field strength of 4 T and has a larger radius, so the entire calorimeter system is housed inside the solenoid.

4.9.2 Magnetic fields for muon spectrometers

One option for the magnetic field for the muon spectrometer is to use magnetized iron. If there is a superconducting solenoid for the tracker, the magnetic flux will return from the solenoid through the iron surrounding the solenoid. In this case, the iron serves multiple purposes: it can be the passive absorber for the hadron calorimeter and act as shielding to remove particles other than muons before they reach the muon chambers, as well as acting as the return ‘yoke’ for the solenoid. The iron is instrumented with tracking chambers (a variety of wire chambers) and the reconstructed muon tracks in these chambers can therefore be used to determine the momenta. The momentum resolution for these tracks is limited by multiple scattering to about 10%. In the CMS approach, the muon spectrometer tracks are linked to the much more precisely measured tracks in the tracker and hence a good muon momentum resolution can be achieved.⁵⁸

In the approach used for ATLAS, the magnetic field for the muon spectrometer is generated by eight large superconducting toroids in the central (‘barrel’) region and eight smaller superconducting coils in each end cap (see Fig. 4.28). The average magnetic field in the tracking volume

⁵⁸See Exercise 4.9 for a discussion of some issues associated with muon triggers for this configuration.

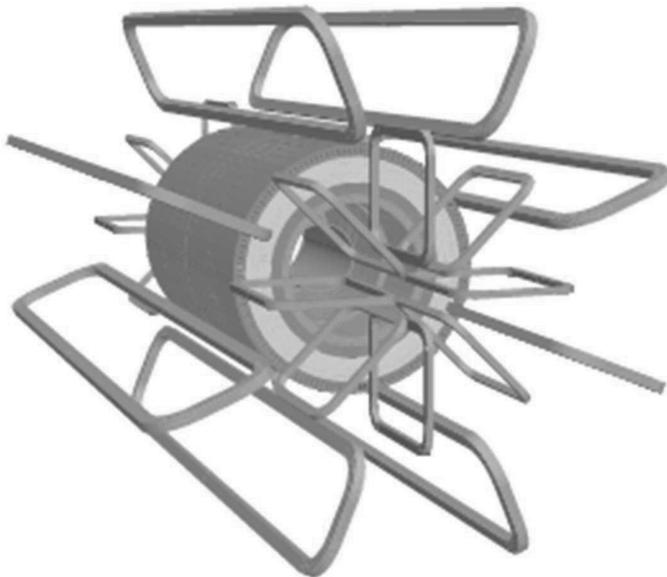


Fig. 4.28 Schematic view of the ATLAS toroid coils [20]. The eight barrel toroid coils with the interleaved end-cap coils are shown. The cylinder shows the return flux for the solenoid. The length is 25.1 m and the outer diameter is 20.1 m.

⁵⁹The total energy stored in the ATLAS magnetic fields is about 1.6 GJ, which is the same magnitude as the kinetic energy in a TGV train with a mass of 385 t travelling at 330 km h⁻¹.

⁶⁰The most precise muon measurement is then obtained by combining the estimates from the tracker and the muon spectrometer.

is in the range $\sim 0.5\text{--}1$ T, but very good resolution is achieved by tracking over a long length $l \sim 5$ m.⁵⁹ Since most of the volume is air, the momentum resolution is not so limited by multiple scattering as for magnetized iron. Another advantage of this field configuration is that it allows reconstruction of precise muon momenta independently of the tracker.⁶⁰

4.10 Trigger

The trigger is an electronic and software system operating in ‘real’ time to reduce the raw data rate to a level that can be permanently stored. The trigger should keep as much of the interesting physics while rejecting the maximal amount of background events. The aim is to bring the rate down from the raw interaction rate to the maximum at which data can be kept in permanent storage, while retaining as large a fraction of the signal events as possible. Traditionally, this rate was typically of the order of a Hz, but advances in computer technology now allow far higher rates. The event rates are very different for different colliders. At e^+e^- colliders, the rates are relatively low, of the order of a Hz, but the rates at hadron colliders have been increasing. At the LHC, there are multiple interactions per bunch crossing (50 ns in 2012 running and 25 ns for the nominal LHC operation) and the trigger reduces this rate to a level of the order of 500 Hz.

Typically, there are three trigger stages or levels:

- In the older generation of experiments, at the first, fastest, level, the selection is based on the timing and the signal level of detector components. The implementation is usually in fast hardware logic operations on outputs from units like comparators and coincidence counters. Detector signals are required to be in coincidence with colliding-beam bunches and to be compatible with tracks and energy deposits of particles coming from a small region where colliding-beam bunches overlap. More sophisticated algorithms are required for the LHC (see Section 4.10.1).
- At the second level, fast processors are used to reject background events, like those coming from cosmic rays or stray accelerator particles in a halo around the beam pipe or from beam particles interacting with molecules and atoms in the residual gas in the beam pipe. At this level, we also need to reject genuine, but uninteresting, physics events produced by colliding-beam particles.
- The third level is often comparable to the offline reconstruction. Farms of computers select signal events to be stored for offline reconstruction and analysis. The main difference between the third-level trigger programs and offline programs is in the use of calibration constants and correction procedures, which need to be obtained or developed separately offline.

4.10.1 LHC triggers

The issue of efficiently triggering on interesting physics events, while maintaining a manageable readout rate, is one of the main challenges for LHC detectors. At design luminosity, the rate of pp collisions is about 1 GHz, and this rate has to be reduced to the order of 500 Hz for data to be stored for subsequent offline analysis. The first-level trigger (L1) uses signals from the full detector, which, given the finite speed of light, makes it impossible to generate a trigger decision from one bunch crossing before the following bunch crossing occurs (25 ns at nominal LHC operation). This apparently insoluble problem is solved with the aid of a ‘pipelined’ system.⁶¹ The data are stored on detector in ‘pipeline’ memory (see Fig. 4.20), while the L1 decision is being made by a custom hardware processor. In such a pipelined processor, one step of the trigger process operates on the data for a particular event in one clock cycle and then the next step is operated in the following clock cycle. The number of allowed steps for such a processor depends on the depth of the pipeline memory in which the data are stored.⁶² As all bunch crossings have genuine pp collisions, it is no longer sufficient to simply reject non-beam backgrounds—the L1 trigger must decide which real events to keep. The L1 trigger uses interesting signatures like high-transverse-momentum electrons by performing hardware sums of the energy deposited in neighbouring cells in the electromagnetic calorimeter. A global L1 trigger decision is made on the basis of several signatures (high-transverse-momentum muon candidates, large missing transverse energy, etc.). This L1 trigger typically reduces the rate to the order of 100 kHz. At this rate, it is now feasible to read out all the data corresponding to triggered bunch crossings⁶³ and for the data to be processed by very large computer farms, which use the full detector granularity to reduce the rate to the required order of 500 Hz for storage.

⁶¹This approach was pioneered by the H1 and ZEUS experiments at DESY for the HERA collider.

⁶²A typical pipeline depth of 132 corresponds to a time of $3.2\mu\text{s}$, which is sufficient to allow the signals to reach the trigger processor, for a trigger decision to be made, and for that decision to be fed back to the electronics on the detector.

⁶³The readout is performed using optical fibre links.

4.11 Examples of detector systems

Now that we have seen the principles behind the design of detector subsystems, we can start to understand how these principles are applied in the designs of real detectors. We first look at collider detectors and then briefly consider neutrino detectors. Dark matter detectors are described in Section 13.7.2.

4.11.1 Collider detectors

We will take the ATLAS and CMS detectors as examples of collider detectors.⁶⁴ The ATLAS detector is illustrated schematically in Fig. 4.29. The tracker is immersed in a 2 T magnetic field and consists of silicon detectors closest to the beam line and a Transition Radiation Tracker (TRT) at larger radius. The silicon detector contains three layers of

⁶⁴We cover some unique aspects of the LHCb detector in Section 10.7.3.

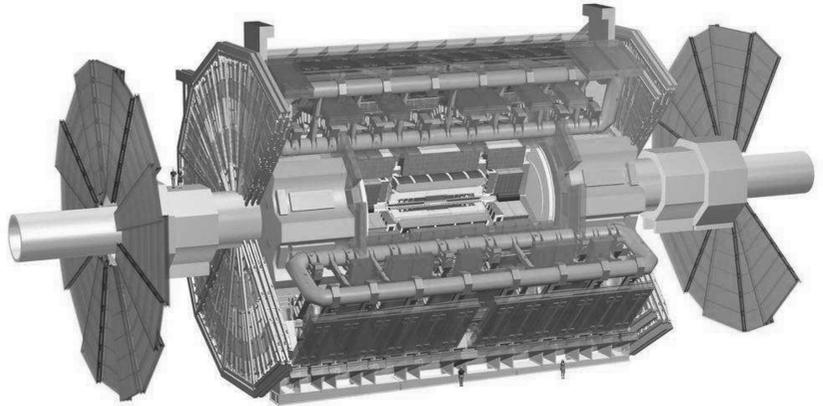


Fig. 4.29 Schematic view of the ATLAS detector [20].

pixels closest to the beam pipe to provide the best resolution for the impact parameter and layers of silicon strips at larger radius. The TRT is made from cylindrical ‘straw’ tubes, with each tube working as an independent cylindrical drift chamber. The tubes are interleaved with Mylar foils to generate transition radiation to enhance electron identification. The electromagnetic calorimeter is based on the liquid-argon accordion calorimeter (see Section 4.7.3). In the central region, the hadronic calorimeter uses an iron–scintillator sandwich design. The light from the scintillators is coupled to the photomultipliers using wavelength-shifting fibres. The novel feature of this design is that the steel absorber plates are rotated by 90° compared with the conventional design in which the plates are perpendicular to the direction of incidence of primary particles. This has the advantage that the calorimeter cells are self-supporting, thus avoiding ‘dead’ material between cells. Although the calorimeter system is not compensating, the fine granularity allows the use of software compensation to improve the resolution. Calorimeters extend up to pseudorapidity $\eta \approx 5$ in order to reconstruct missing transverse momentum (see Chapter 8). The muon spectrometer uses the toroidal coils discussed in Section 4.9.2. In the central barrel region, the muon tracks are measured using detectors based on drift tubes. However, the signals are too slow to participate in the first-level trigger (see Section 4.10) and therefore faster but lower-resolution detectors are also used.

A schematic view of the CMS detector is shown in Fig. 4.30. There is a very large all-silicon tracker consisting of three layers of pixel detector and 10 layers of strip detectors immersed in the 4T solenoidal magnetic field, which provides very good momentum resolution for charged particles. The electromagnetic calorimeter uses PbWO_4 crystals (see Section 4.7.2). The hadronic calorimeter uses a brass–scintillator sandwich calorimeter. As with ATLAS, forward calorimeters extend the coverage to close to the beam pipe. The muon chambers are interleaved with the return yoke of the solenoid. They are used for the first-level muon trigger, but high-precision measurements of muon momentum are made in the tracker.

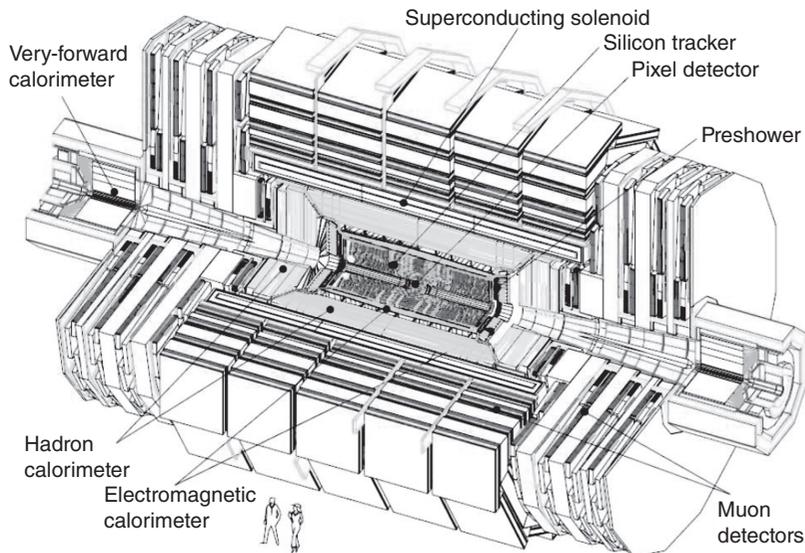


Fig. 4.30 Schematic view of the CMS detector [62].

4.11.2 Neutrino detectors

Optimization of neutrino detectors is very different to that of collider detectors because the very small cross sections mean that very massive detectors are needed to allow useful event rates to be obtained.⁶⁵ Given the sizes involved, we are obliged to use cheaper detector technologies than at hadron colliders. The requirements depend on the neutrino energies. For an accelerator neutrino experiment, a typical requirement is to have a very large target mass and be able to measure the following:

- **Electrons:** We need to measure electrons from neutral-current scattering (see Chapter 7) off electrons $\nu_x e^- \rightarrow \nu_x e^-$ (where ν_x is any flavour of neutrino), or from similar processes with scattering on the nuclei.
- **Muons:** We have muons from charged-current interactions $\nu_\mu N \rightarrow \mu^- N' X$, where N and N' are the target and scattered nuclei, respectively, and X represents any hadrons produced in the interaction.
- **Hadrons:** For neutral-current scattering off nuclei, the only particles we can measure are the outgoing hadrons. Measurements of produced hadrons also improve the determination of the event kinematics for charged-current interactions.

In general, we can use calorimeters to measure electrons and hadrons. If the passive absorber plates are made from magnetized iron and we instrument the gaps between absorbers with some tracking detector, we can determine the tracks caused by muons. We can then identify muons as particles that penetrate deeper into the detector than hadrons and at the same time we can estimate the momentum by measuring the curvature of the tracks. We will see how these principles are applied in practice in the MINOS far detector in Chapter 11.

⁶⁵This is particularly true for neutrino detectors in laboratory oscillation experiments, in which we need a detector far from the neutrino source to study oscillations (see Chapter 11).

Chapter summary

- The physics of interactions of high-energy particles in matter has been reviewed.
- The basic detector physics of how a signal is generated by charged particles has been explained.
- A brief summary of how different types of scintillators work has been given.
- The basic concepts of trigger systems have been explained.
- Different detector systems have been discussed, and it has been described how they are combined into a general purpose detector.
- Further case studies of real particle physics detectors are given in other chapters.

Further reading

- Particle Data Group (2014). Review of Particle Physics. *Chin. Phys. C*, **38**, 090001. In the section ‘Experimental methods and colliders’, the review article ‘Passage of particles through matter’ gives a thorough discussion. The review article ‘Particle detectors at accelerators’ gives a more advanced and thorough discussion than is given in the present chapter.
- Green D. (Ed.) (2010). *At the Leading Edge: The ATLAS and CMS LHC Experiments*. World Scientific. This is a collection of advanced review articles on different aspects of the detectors.
- Grupen, C. and Shwartz, B. (2008). *Particle Detectors* (2nd edn). Cambridge University Press. This gives a very comprehensive description of many detector technologies.
- Kleinknecht, K. (1998). *Detectors for Particle Radiation* (2nd edn). Cambridge University Press. This gives a short and clear introduction to detector physics.
- Blum, W., Riegler, W., and Rolandi, L. (2008). *Particle Detection with Drift Chambers* (2nd edn). Springer. This is the definitive advanced textbook on this subject.
- Spieler, H. (2005). *Semiconductor Detector Systems*. Oxford University Press. This is a very good advanced textbook on silicon detectors and the associated electronics.

Exercises

- (4.1) Starting from eqn 9.12 in Chapter 9 and using a change of variable, derive eqn 4.1.
- (4.2) Consider elastic scattering of a heavy particle of mass M with speed β on a stationary electron.
- (a) Let the kinetic energy of the scattered electron be T in the frame in which the electron was initially at rest. Show that the 4-momentum transfer evaluated in this frame is $Q^2 = 2m_e T$.
- (b) Assuming $m/M \ll 1$ and $\gamma m/M \ll 1$, show that the maximum kinetic energy of the electron after the scattering in the lab frame is $T_{\max} = 2\gamma^2 \beta^2 m_e$.
- Hint:* Consider the problem in the CMS and then use a Lorentz transformation from the CMS to the lab.

- (4.3) Calculate the Fourier coefficients for the periodic function defined by $f(z) = z/L$ for $0 < z < L$ and $f(z) = 0$ for $-L < z < 0$ (the function repeats periodically). Use this result to derive eqn 4.17 starting from eqn 4.16.
- (4.4) (a) A very crude model of the initial development of an electromagnetic shower is that a high-energy electron or positron of energy E_0 undergoes a bremsstrahlung process after a distance L (one radiation length) and loses half of its energy to a secondary photon, or that a high-energy photon initiates a pair production process after travelling a distance L , splitting its energy equally between the two secondary particles. These processes continue until the photons and charged particles each have an energy less than the critical energy $E_c (\ll E_0)$, at which point the multiplication ceases. Develop this model and answer the following questions both for an incident electron of energy E_0 and for an incident photon of the same energy:
- How many photons and charged particles will there be after N radiation lengths?
 - What is the energy of each particle in the shower after N radiation lengths?
 - What is the depth (in units of L) at which the number of particles in the shower is a maximum, and what is the number of particles at maximum?
- (b) Compute the depth and the number of particles when multiplication ceases for a 4 GeV electron entering lead glass ($L = 2.5$ cm, $E_c = 10$ MeV).
- (c) Explaining any assumptions you make, how would the resolution of an electromagnetic calorimeter scale with the energy of the incident electron, E ?
- (4.5) Calculate the direction of Čerenkov radiation with respect to the direction of motion of fast charged particles in water. The refractive index of water is 1.33.
- (4.6) Calculate the threshold energy above which electrons and muons emit Čerenkov radiation in water (refractive index = 1.33). What consequences does this have for the measurement of
- the solar neutrino flux?
 - the flavour ratio of atmospheric neutrinos?
- (4.7) A very simple model of a high-precision silicon ‘micro-vertex detector’ (MVD) consists of two concentric cylindrical layers surrounding the beam line. The first layer is at radius $R_0 = 5$ cm, and the separation between the first and second layers is $L = 2$ cm. The intrinsic measurement resolution of a hit is $\sigma = 10$ μm in the $R\phi$ direction (roughly orthogonal to the trajectory of a particle with large transverse momentum).
- Show that (neglecting multiple scattering) the uncertainty in the impact parameter (distance of closest approach to the beam line in the plane perpendicular to the beam), σ_a , is given by
- $$\sigma_a = \frac{\sigma}{L} \sqrt{(R_0 + L)^2 + R_0^2}$$
- and calculate it for the parameter values given above. How does σ_a change if (i) L is doubled; (ii) R_0 is increased to 8 cm? What factors limit the ability to decrease R or increase L . Assume that each layer has a thickness of 2% of a radiation length. How does multiple scattering affect the impact parameter resolution? For what momentum would the uncertainty in the impact parameter from measurement error be equal to that from multiple scattering?
- (4.8) Consider a cylindrical detector immersed in a uniform solenoidal magnetic field B . Let R be the radius of curvature of a track in the plane transverse to the beam line (measured in metres). Show that the transverse momentum $p_T = 0.3BR$. A very simplified model for the resolution of a tracker assumes that the track is precisely located at the start and end of the trajectory but there is a measurement error in the transverse plane of σ_s at a radius of half the outer radius of the tracker ($L/2$). Using this model, determine the transverse momentum resolution as a function of p_T . For such a detector with $B = 4$ T, $L = 1$ m, and $\sigma_s = 10$ μm , estimate the largest value of p_T that could be measured with an error less than one-third of the value.
- (4.9) Consider a solenoid providing a uniform magnetic field $\mathbf{B} = B_1 \hat{\mathbf{z}}$ for a radius $0 < R < R_1$. All the flux returns through a return yoke such that $\mathbf{B} = B_2 \hat{\mathbf{z}}$ for a radius $R_1 < R < R_2$.
- Show that $\int_0^{R_2} B(r) dr = 0$.
 - What is the force on a charged particle moving with a velocity v in the (x, y) plane? Hence find the torque on the charged particle.

- (c) Now consider the trajectory of a muon created on the axis of the solenoid ($r = 0$). Combining the results of (a) and (b), show that there is no net change in angular momentum of the muon during its trajectory from $r = 0$ to $r = R_2$. Explain why this means that the trajectory of the muon after it exits the return yoke (i.e. at $r = R_2$) points back to the axis
- (d) These calculations have ignored multiple scattering—how would this change the result qualitatively?
- (e) Discuss the implications for the measurement of muon momenta in this geometry.
- (4.10) We wish to measure a charged particle with momentum transverse to the beam line of $p_T = 500 \text{ GeV}$ in a tracking detector immersed in a solenoidal field $\mathbf{B} = 2 \text{ T}$. If we require a momentum resolution $\sigma(1/p_T)/(1/p_T) = 0.3$, estimate the spatial resolution required for the sagitta measurement. Discuss which detector technology would be appropriate.
- (4.11) Consider a cylindrical drift chamber with a radius of 4 mm, operated at a voltage of 2 kV. If the positive ions have a mobility $\mu = 1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, calculate the maximum drift time. How long does it take to accumulate 50% of the full signal?
- (4.12) Consider an MWPC with anode spacing d . Consider the coordinate x in the plane of the anode wires. Calculate the root mean square difference in x between the location of a track and the nearest wire, and hence justify the claim that the resolution is $d/\sqrt{12}$.
- (4.13) Consider an n -doped semiconductor with carrier densities n and p of electrons and holes, respectively. Assuming that $n \gg p$, show that the electrical conductivity $\sigma = ne\mu_e$ where μ_e is the electron mobility. A pn silicon microstrip detector (see Fig. 4.19), with the resistivity of the silicon $\rho = 10 \text{ k}\Omega \text{ cm}$, has a thickness $w = 300 \mu\text{m}$. The relative permittivity of Si is $\epsilon_r = 11.6$ and $\mu_e = 1350 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Determine the bias voltage required to fully deplete the detector, $V_{\text{depletion}}$.
- If an electron–hole pair is created at a distance x from the p -type electrode, calculate the drift time of the hole in terms of the mobility of the holes, μ_h . For silicon, with $\mu_h = 480 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, determine the charge collection times for holes created at depths $x = 0.5w$ and $x = 0.9w$. If the detector were operated at a bias voltage $V = 2V_{\text{depletion}}$, how would the charge collection times change? Hence discuss the advantages of operating the detector at a voltage greater than the depletion voltage. What limits the detector voltage that can be applied in practice?
- (4.14) The leakage current in a silicon detector is a source of noise. If the leakage current in one channel is I_{leak} and the signal is integrated over a time T , make a simple estimate of $\sigma(Q_{\text{leak}})$, the contribution of the leakage current to the noise on the charge signal. For a typical LHC silicon detector, $T \sim 25 \text{ ns}$ (the bunch spacing). Estimate $\sigma(Q_{\text{leak}})$ for two cases: (a) $I_{\text{leak}} = 1 \text{ nA}$ (typical for an un-irradiated strip detector) and (b) $I_{\text{leak}} = 1 \mu\text{A}$ (typical for a heavily irradiated strip detector). Compare these noise values with the signal expected from a $300 \mu\text{m}$ -thick silicon detector. Design a simple filter circuit to minimize the leakage current noise while keeping as much as possible of the signal. Suggest an approximate value for the cut-off frequency of your filter for an LHC microstrip detector.
- (4.15) Consider a silicon microstrip detector with p -doped implants (strips) in n -doped bulk silicon (see Fig. 4.19). Make a rough sketch of the ‘weighting’ field (see Section 4.4.1) in the region around one strip and indicate on it the region in which the weighting field will be large. A charged particle crosses such a detector in a direction perpendicular to the plane of the silicon and creates electron–hole pairs uniformly along its trajectory. For a reverse-biased detector, which way will the electrons (holes) drift? By combining the above considerations, show that the resulting signal will be dominated by the motion of holes, rather than electrons.

Static quark model

5

The static quark model of hadrons is central to the understanding of the pattern of hadronic masses and quantum numbers. Originally devised with three flavours of quarks (u , d , s), it was extended to include the heavy quarks—first charm after the ‘1974 revolution’ when the J/ψ was discovered and shown to be a $q\bar{q}$ state and then beauty some years later. The top quark was long anticipated on the basis of quark–lepton ‘generation’ symmetry after the tau lepton was discovered in 1975,¹ but proved to be enormously heavy when it was finally teased out of the data by the CDF and D0 experiments at the Tevatron. The pattern of quark masses is one of the big unsolved problems of particle physics. In this chapter, we are concerned primarily with how the quark model helps our understanding of the phenomenology of mesons and baryons.

The chapter starts with a reminder of the 2-component spin- $\frac{1}{2}$ algebra and its connection with the $SU(2)$ group. Then, after a brief account of hadronic isospin (based on $SU(2)$), we explain how this approximate symmetry is extended, by including strangeness, to the flavour $SU(3)$ of the static quark model.²

5.1 Spin $\frac{1}{2}$	115
5.2 The quark model of hadrons	118
5.3 Heavy quarks	128
5.4 Exotic hadrons	137
Chapter summary	139
Further reading	139
Exercises	139

¹The discovery of the associated neutrino τ_ν is discussed in Chapter 8.

²The definition and properties of the $SU(n)$ groups are given in Chapter 2.

5.1 Spin $\frac{1}{2}$

For a half-integer spin fermion, the eigenfunctions for spin up ($m = +\frac{1}{2}$) and down ($m = -\frac{1}{2}$) are

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The raising and lowering operators, defined as (see Chapter 2)

$$s_{\pm} = s_x \pm is_y$$

are, by inspection,

$$s_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad s_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

This is easily demonstrated—for example

$$s_+ \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$s_+ \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Working backwards, we get expressions for s_x and s_y :

$$s_x = \frac{1}{2}(s_+ + s_-), \quad s_y = \frac{1}{2}i(s_- - s_+)$$

and s_z is derived from the requirement that

$$s_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad s_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

We can now write the spin- $\frac{1}{2}$ defining expressions in terms of the Pauli matrices (note the factor $\frac{1}{2}$):

$$s_x = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} = \frac{1}{2}\sigma_x$$

$$s_y = \begin{pmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{pmatrix} = \frac{1}{2}\sigma_y$$

$$s_z = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} = \frac{1}{2}\sigma_z$$

The spin operator algebra can be summarized as

$$\left[\frac{1}{2}\sigma_i, \frac{1}{2}\sigma_j \right] = \epsilon_{ijk} \frac{1}{2}i\sigma_k$$

The link with SU(2) is clear:

- The Pauli spin matrices are Hermitian and traceless.
- There are three of them, as expected.
- The set of unitary matrices $U_i(\theta_i) = e^{-\frac{1}{2}i\sigma_i\theta_i}$ form the fundamental (irreducible)³ representation of SU(2).
- The pattern of combinations is predicted from representation theory.

³See Chapter 2 for more on groups and SU(2).

5.1.1 Combining two spin- $\frac{1}{2}$ particles

Two spin- $\frac{1}{2}$ particles can combine to form four possible spin states. The total spin can be either $s = 1, s_z \in \{-1, 0, 1\}$ or $s = 0, s_z = 0$. We have

$$|1, 1\rangle = \left| \frac{1}{2}, \frac{1}{2} \right\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = |\uparrow, \uparrow\rangle$$

$$|1, 0\rangle = \sqrt{\frac{1}{2}} \left| \frac{1}{2}, \frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \sqrt{\frac{1}{2}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} (|\uparrow, \downarrow\rangle + |\downarrow, \uparrow\rangle)$$

$$|1, -1\rangle = \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = |\downarrow, \downarrow\rangle$$

$$|0, 0\rangle = \sqrt{\frac{1}{2}} \left| \frac{1}{2}, \frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \sqrt{\frac{1}{2}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} (|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)$$

The $s = 1$ triplet is symmetric under the interchange of particles. They are deduced from the spin-raising/lowering operators and the Clebsch-Gordan coefficients.⁴

⁴See Chapter 2.

The $s = 0, s_z = 0$ singlet state is found by requiring it to be orthogonal to the $s = 1, s_z = 0$ state. Note that the $s = 1$ states are symmetric under the interchange of particles 1 and 2, whereas the $s = 0$ state is antisymmetric. This illustrates what is meant by the representation notation

$$2 \otimes 2 = \underbrace{3}_{\text{sym}} \oplus \underbrace{1}_{\text{asym}}$$

where ‘sym’ and ‘asym’ stand respectively for symmetric and antisymmetric combinations of the spin states of the two particles.

5.1.2 Combining three spin- $\frac{1}{2}$ particles

We will now show that the multiplicity of states is given by $2 \otimes 2 \otimes 2 = 4 \oplus 2 \oplus 2$.⁵ We start by adding a spin-up particle to the $|1, 1\rangle$ to obtain the highest possible (maximally stretched) state $|\frac{3}{2}, +\frac{3}{2}\rangle$, then apply angular-momentum-lowering operators to ‘step down’ from there:

⁵The multiplicities must be even for half-integer angular momentum.

$$\begin{aligned} \left| \frac{3}{2}, \frac{3}{2} \right\rangle &= | \uparrow \uparrow \uparrow \rangle \\ \left| \frac{3}{2}, \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{3}} (| \downarrow \uparrow \uparrow \rangle + | \uparrow \downarrow \uparrow \rangle + | \uparrow \uparrow \downarrow \rangle) \\ \left| \frac{3}{2}, -\frac{1}{2} \right\rangle &= \sqrt{\frac{1}{3}} (| \downarrow \downarrow \uparrow \rangle + | \uparrow \downarrow \downarrow \rangle + | \downarrow \uparrow \downarrow \rangle) \\ \left| \frac{3}{2}, -\frac{3}{2} \right\rangle &= | \downarrow \downarrow \downarrow \rangle \end{aligned}$$

Two more states come from adding the third particle to the triplet. Requiring orthogonality to the $S = \frac{3}{2}, S_z = +\frac{1}{2}$ states, two $S = \frac{1}{2}$ states occur:

$$\begin{aligned} \left| \frac{1}{2}, \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{6}} (2| \uparrow \uparrow \downarrow \rangle - | \downarrow \uparrow \uparrow \rangle - | \uparrow \downarrow \uparrow \rangle) \\ \left| \frac{1}{2}, -\frac{1}{2} \right\rangle &= -\sqrt{\frac{1}{6}} (2| \downarrow \downarrow \uparrow \rangle - | \uparrow \downarrow \downarrow \rangle - | \downarrow \uparrow \downarrow \rangle) \end{aligned}$$

These states are of mixed symmetry.

Finally, two states are derived from adding the third particle to the singlet:

$$\begin{aligned} \left| \frac{1}{2}, \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{2}} (| \uparrow \downarrow \uparrow \rangle - | \downarrow \uparrow \uparrow \rangle) \\ \left| \frac{1}{2}, -\frac{1}{2} \right\rangle &= -\sqrt{\frac{1}{2}} (| \uparrow \downarrow \downarrow \rangle - | \downarrow \downarrow \uparrow \rangle) \end{aligned}$$

which are antisymmetric combinations under the exchange of particles $1 \leftrightarrow 2$. The pattern of combinations is indeed

$$2 \otimes (2 \otimes 2) = \underbrace{4}_{\text{sym}} \oplus \underbrace{2}_{\text{mix}_S} \oplus \underbrace{2}_{\text{mix}_A}$$

where ‘ $\text{mix}_{(A)S}$ ’ means a state that is a mixture of symmetric and antisymmetric parts but becomes purely (anti)symmetric under exchange of the first two particles.

5.2 The quark model of hadrons

Quarks are fundamental, point-like constituents of matter that carry either $\frac{1}{3}$ or $\frac{2}{3}$ electric charge. They exist in $q\bar{q}$ or qqq bound states—the hadrons—held together by the strong force. The strong force is mediated by gluons, which couple with equal strength to all quarks. The residual strong force holds together the nucleus in analogy to the van der Waals force between neutral atoms.

Strong-force bound states can be organized by a classification that predates the invention of the quark model; it arose from similarities among the hadrons observed in the 1960s, particularly from bubble chamber experiments. The observed patterns of the quantum numbers (mass, spin, parity, isospin, and strangeness) were crucial in the development of the quark model.

Baryons

- Baryons consist of three quarks or three antiquarks.
- They are fermions (with spins $J = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$).
- One baryon, the proton, is stable. The neutron is almost stable, forming stable bound states with the proton in many atomic nuclei. However, a free neutron is not stable, decaying to a proton, an electron, and an antineutrino with a mean lifetime of 885.7 ± 0.8 s.
- Baryon number conservation was invented to explain the non-observation of the lightest baryons decaying to purely mesonic or leptonic final states. In the quark model, this translates into quark number conservation.

Mesons

- Mesons are quark–antiquark pairs
- They are bosons, since they have integer spin ($J = 0, 1, 2, \dots$)
- All mesons are unstable. The most stable is the pion, with $\tau(\pi^\pm) = 26$ ns ($c\tau \approx 8$ m).

Although the quark model is now taken for granted, it is useful to think about how physicists used the data to devise it and why the concepts matched so beautifully the experimental observations.

5.2.1 Isospin

Originally, physicists were intrigued by the similarity in mass and properties of the proton and neutron. Inspired by spin- $\frac{1}{2}$ doublets, it was postulated they were the same particle⁶ but with a different projection of the third component of a new angular-momentum-like quantum number—*isospin*. In fact, this is not the case, but it is still a useful approximate symmetry for low-energy hadronic physics, where perturbative QCD is not valid.

⁶The mass difference $m_n - m_p$ is of the order of an electromagnetic correction.

The key points are as follows:

- $|u\rangle$ and $|d\rangle$ form an $SU(2)$ doublet like spin $\frac{1}{2}$:

$$I_3|u\rangle = \frac{1}{2}|u\rangle, \quad I_3|d\rangle = -\frac{1}{2}|d\rangle$$

- As with normal (angular momentum) spin, the raising and lowering operators are

$$I_+|d\rangle = |u\rangle, \quad I_-|u\rangle = |d\rangle, \quad I_-|d\rangle = 0$$

- Although the formalism is the same, isospin has **nothing** to do with angular momentum.
- Isospin is an approximately conserved quantity in strong interactions (it would be exact if the masses of the u and d quarks were identical).
- The validity of isospin is now understood to follow from two fundamental assumptions:
 - (1) the approximate degeneracy in mass of the u and d quarks;
 - (2) u and d have an identical strong coupling to the gluon.

When dealing with antiquarks, one must be careful in applying charge conjugation to the quark wavefunctions. We choose a convention that allows Clebsch–Gordan coefficients to be applied in the same manner as with quarks, although this introduces a somewhat confusing minus sign:⁷

$$C|u\rangle = -|\bar{u}\rangle, \quad C|d\rangle = |\bar{d}\rangle$$

⁷We follow the convention defined originally for atomic physics by Condon and Shortley in their famous book *The Theory of Atomic Spectra* [68].

where C is the charge-conjugation operator. The raising and lowering operators act on the antiquarks as follows:

$$I_-|\bar{d}\rangle = -|\bar{u}\rangle, \quad I_+|\bar{u}\rangle = -|\bar{d}\rangle, \quad I_+|\bar{d}\rangle = 0, \quad I_-|\bar{u}\rangle = 0$$

Consider next the $SU(2)$ isospin combinations of $2 \otimes \bar{2} = 3 \oplus 1$, where the Clebsch–Gordan coefficients are taken from the $\frac{1}{2} \otimes \frac{1}{2}$ table: a symmetric singlet state

$$|I = 0, I_3 = 0\rangle = \sqrt{\frac{1}{2}}(|d\bar{d}\rangle + |u\bar{u}\rangle)$$

and an antisymmetric triplet state

$$\begin{aligned} |I = 1, I_3 = 1\rangle &= |u\bar{d}\rangle &&= |\pi^+\rangle \\ |I = 1, I_3 = 0\rangle &= \sqrt{\frac{1}{2}}(|d\bar{d}\rangle - |u\bar{u}\rangle) &&= |\pi^0\rangle \\ |I = 1, I_3 = -1\rangle &= -|\bar{u}d\rangle &&= |\pi^-\rangle \end{aligned}$$

⁸A mass difference again of a magnitude compatible with an electromagnetic effect.

The triplet of pions is almost degenerate in mass: $m(\pi^\pm) = 140 \text{ MeV}$, $m(\pi^0) = 135 \text{ MeV}$.⁸

Isospin is also useful for baryons. As with spin, we expect a symmetric quadruplet and two mixed-symmetry isospin doublets: $2 \otimes (2 \otimes 2) = 4 \oplus 2 \oplus 2$, which we can write out explicitly:

$$\begin{aligned} \left| I = \frac{3}{2}, I_3 = \frac{3}{2} \right\rangle &= |uuu\rangle \\ \left| I = \frac{3}{2}, I_3 = \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{3}}(|duu\rangle + |udu\rangle + |uud\rangle) \\ \left| I = \frac{3}{2}, I_3 = -\frac{1}{2} \right\rangle &= \sqrt{\frac{1}{3}}(|ddu\rangle + |udd\rangle + |dud\rangle) \\ \left| I = \frac{3}{2}, I_3 = -\frac{3}{2} \right\rangle &= |ddd\rangle \\ \left| I = \frac{1}{2}, I_3 = \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{6}}(2|uud\rangle - |duu\rangle - |udu\rangle) \\ \left| I = \frac{1}{2}, I_3 = -\frac{1}{2} \right\rangle &= -\sqrt{\frac{1}{6}}(2|ddu\rangle - |udd\rangle - |dud\rangle) \\ \left| I = \frac{1}{2}, I_3 = \frac{1}{2} \right\rangle &= \sqrt{\frac{1}{2}}(|udu\rangle - |duu\rangle) \\ \left| I = \frac{1}{2}, I_3 = -\frac{1}{2} \right\rangle &= -\sqrt{\frac{1}{2}}(|udd\rangle - |dud\rangle) \end{aligned}$$

An example of isospin analysis: Δ decays

⁹Higher-mass baryon states are not accessible.

- Baryon number conservation requires the Δ to decay to p or n .⁹
- Strong-interaction decays are favoured by a factor $O(10^4)$ over electromagnetic decays.
- As the strong interaction dominates, we can use isospin to understand *relative* rates using

$$\begin{aligned} |\Delta^+\rangle &= \left| I = \frac{3}{2}, I_3 = \frac{1}{2} \right\rangle, & |p\rangle &= \left| \frac{1}{2}, \frac{1}{2} \right\rangle, & |n\rangle &= \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \\ |\pi^+\rangle &= |1, 1\rangle, & |\pi^0\rangle &= |1, 0\rangle, & |\pi^-\rangle &= |1, -1\rangle \end{aligned}$$

- Using Clebsch–Gordan coefficients, we expand the $I = \frac{3}{2}$ in products of $I = \frac{1}{2}$ and $I = 1$ states:

$$\begin{aligned}\Delta^+ &= \left| \frac{3}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{3}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle |1, 1\rangle + \sqrt{\frac{2}{3}} \left| \frac{1}{2}, \frac{1}{2} \right\rangle |1, 0\rangle \\ &= \sqrt{\frac{1}{3}} |n\rangle |\pi^+\rangle + \sqrt{\frac{2}{3}} |p\rangle |\pi^0\rangle\end{aligned}$$

- Deducing branching ratios, we have

$$\frac{\text{BR}(\Delta^+ \rightarrow \pi^0 p)}{\text{BR}(\Delta^+ \rightarrow \pi^+ n)} = \frac{|\langle \pi^0 p | \Delta^+ \rangle|^2}{|\langle \pi^+ n | \Delta^+ \rangle|^2} = \frac{\left| \sqrt{\frac{2}{3}} \right|^2}{\left| \sqrt{\frac{1}{3}} \right|^2} = 2$$

- Similarly, we can estimate the relative Δ cross sections for formation in πp scattering at $\sqrt{s} \sim m(\Delta)$:

$$\frac{\sigma(\pi^- p)}{\sigma(\pi^+ p)} = \frac{|\langle \Delta^0 | \pi^- p \rangle|^2}{|\langle \Delta^{++} | \pi^+ p \rangle|^2} = \frac{1}{3}$$

- If we assume that the cross section on resonance is dominated by the Δ , the data are in reasonable agreement with the isospin assignment: $\sigma(\pi^- p) \approx 70$ mb, $\sigma(\pi^+ p) \approx 200$ mb. See Fig. 5.1.
- Remember that isospin analysis is not exact—its usefulness arises from the *approximate* degeneracy of the $|u\rangle$ and $|d\rangle$ masses.

Although useful, isospin could not account for long-lived particles, originally labelled ‘V particles’, that were first observed in 1947 (in Manchester) with a mass ~ 500 times that of the electron (see Fig. 5.2). These new particles—known as ‘strange’ particles—were assigned a new quantum number and SU(2) had to be expanded.

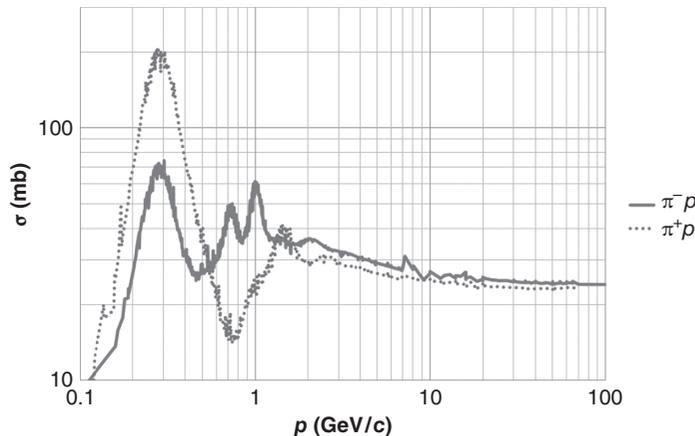


Fig. 5.1 Total cross sections for $\pi^+ p$ (dotted line) and $\pi^- p$ (solid line) as functions of the pion beam momentum in GeV/c. From the PDG [114].

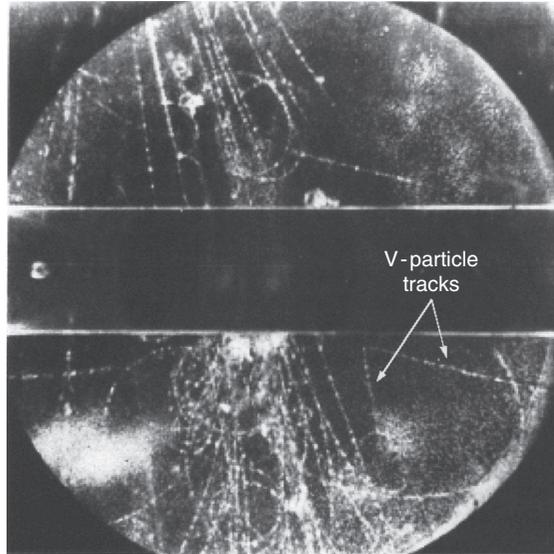


Fig. 5.2 One of the early ‘V particles’ observed in 1947 in a cloud chamber exposed to cosmic rays by Rochester and Butler [123]—then working in Blackett’s group at Manchester University.

¹⁰Generically known now as a *flavour* quantum number.

5.2.2 Strangeness and expansion to SU(3)

With isospin in hand to describe *up*-ness and *down*-ness, strangeness is postulated to be a third quantum number that a hadron may possess.¹⁰ For historical reasons, convention dictates that $s = -1$ for $|s\rangle$ and $s = +1$ for $|\bar{s}\rangle$. Assuming that SU(3) is valid, we expect $3^2 - 1 = 8$ fundamental operators.

The fundamental representation of SU(3) comprises eight 3×3 matrices:

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \sqrt{\frac{1}{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \end{aligned}$$

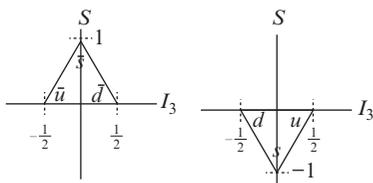


Fig. 5.3 Fundamental SU(3) flavour representations for the (u, d, s) and $(\bar{u}, \bar{d}, \bar{s})$ triplets, as functions of the third component of isospin I_3 (horizontal) and strangeness S (vertical).

SU(3) has three SU(2) groups embedded within it. In addition to isospin (*I-spin*), there are *U-spin* and *V-spin*, which are the doublets of $\begin{pmatrix} s \\ d \end{pmatrix}$ and $\begin{pmatrix} s \\ u \end{pmatrix}$, respectively. The fundamental representation for the quark and antiquark triplets is shown in Fig. 5.3.

The raising and lowering operators are

$$\begin{aligned} I_+|d\rangle &= |u\rangle, & I_-|u\rangle &= |d\rangle, & I_+|\bar{u}\rangle &= -|\bar{d}\rangle, & I_-|\bar{d}\rangle &= -|\bar{u}\rangle \\ U_+|s\rangle &= |d\rangle, & U_-|d\rangle &= |s\rangle, & U_+|\bar{d}\rangle &= -|\bar{s}\rangle, & U_-|\bar{s}\rangle &= -|\bar{d}\rangle \\ V_+|s\rangle &= |u\rangle, & V_-|u\rangle &= |s\rangle, & V_+|\bar{u}\rangle &= -|\bar{s}\rangle, & V_-|\bar{s}\rangle &= -|\bar{u}\rangle \end{aligned}$$

Any other combination is zero. The Condon–Shortley convention is used for the antiquarks, which gives rise to the ‘extra’ minus signs. Only two of the three are needed to describe and navigate through a multiplet, I -spin and U -spin being most commonly used.

5.2.3 Mesons

We now have the tools to extend beyond the π^+, π^0, π^- states of $SU(2)$ $q\bar{q}$ combinations and identify all the $SU(3)$ $q\bar{q}$ states. We start with the simplest case of $J = 0$ pseudoscalar mesons. From $SU(3)$, we expect

$$3 \otimes \bar{3} = 9 = \underbrace{1}_{\text{sym}} \oplus \underbrace{8}_{\text{mix}}$$

where ‘sym’ and ‘mix’ stand for symmetric and mixed symmetry, respectively. The six states with non-zero I and non-zero S have unambiguous quark content. The final three, with $I = S = 0$, need a little more thought:

- The symmetric singlet is ‘obvious’ by inspecting the perfect symmetry of the flavour wavefunction. It is now known as the η' :

$$|\eta'\rangle = \sqrt{\frac{1}{6}} (|u\bar{u}\rangle + |d\bar{d}\rangle + |s\bar{s}\rangle)$$

The remaining two must be part of the octet of mixed symmetry. The first step is to start with the ‘outer’ states of the octet and apply the flavour-lowering operators to reach the centre:

$$\begin{aligned} I_-|u\bar{d}\rangle &= |d\bar{d}\rangle - |u\bar{u}\rangle \\ U_-|d\bar{s}\rangle &= |s\bar{s}\rangle - |d\bar{d}\rangle \\ V_-|u\bar{s}\rangle &= |s\bar{s}\rangle - |u\bar{u}\rangle \end{aligned}$$

However, only two of these three equations are independent. So we proceed as follows:

- We choose one to be the well-established π^0 , the isospin-triplet partner of the π^+, π^- :

$$|\pi^0\rangle = \sqrt{\frac{1}{2}} (|d\bar{d}\rangle - |u\bar{u}\rangle)$$

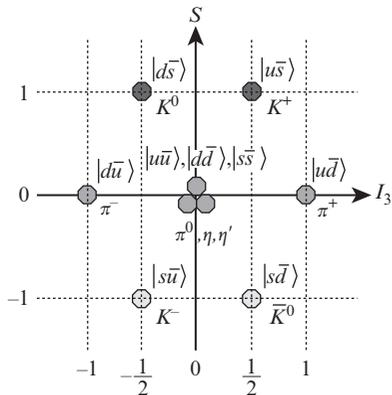


Fig. 5.4 SU(3) pseudoscalar meson nonet states as functions of I_3 and S .

¹¹In fact, it is also slightly broken in the pseudoscalar case.

- This leaves the last member of the octet to be deduced by requiring that its flavour wavefunction be orthogonal to the $|\pi^0\rangle$:

$$\begin{aligned} |\eta\rangle &= \alpha(|s\bar{s}\rangle - |u\bar{u}\rangle) + \beta(|s\bar{s}\rangle - |d\bar{d}\rangle) \\ &= \sqrt{\frac{1}{6}}(|u\bar{u}\rangle + |d\bar{d}\rangle - 2|s\bar{s}\rangle) \end{aligned}$$

where the constants are derived using $\langle\pi^0|\eta\rangle = 0$ and $\langle\eta|\eta\rangle = 1$.

The nonet of pseudoscalar meson states are plotted as functions of I_3 and S in Fig. 5.4 and their properties are listed in Table 5.1.

The $J = 1$ vector mesons are also well-established states. As we might expect, they exhibit the same pattern of states as the $J = 0$ mesons: $3 \otimes 3 = 8 \oplus 1$.

The nonet of vector meson states are plotted as functions of I_3 and S in Fig. 5.5 and their properties are listed in Table 5.2.

The notable difference between the pseudoscalar and vector mesons is that with the latter, for the $I_3 = 0, S = 0$ states, SU(3) is not exact and ‘octet–singlet’ mixing occurs.¹¹ Experimentally, one state (ϕ) is observed to decay largely to kaons, and the other (ω) nearly always to pions. We therefore assume that the states with $I_3 = 0, S = 0$ are maximally mixed,

I	I_3	S	Meson	Composition	Decay	Mass (MeV)
1	1	0	π^+	$ u\bar{d}\rangle$	$\pi^+ \rightarrow \mu^+ \nu_\mu$	140
1	-1	0	π^-	$ \bar{u}d\rangle$	$\pi^- \rightarrow \mu^- \bar{\nu}_\mu$	140
1	0	0	π^0	$\frac{\sqrt{1}}{2}(d\bar{d}\rangle - u\bar{u}\rangle)$	$\pi^0 \rightarrow \gamma\gamma$	135
$\frac{1}{2}$	$\frac{1}{2}$	1	K^+	$ u\bar{s}\rangle$	$K^+ \rightarrow \mu^+ \nu_\mu$	494
$\frac{1}{2}$	$-\frac{1}{2}$	1	K^0	$ d\bar{s}\rangle$	$K^0 \rightarrow \pi\pi(\pi)$	498
$\frac{1}{2}$	$-\frac{1}{2}$	-1	K^-	$ \bar{u}s\rangle$	$K^- \rightarrow \mu^- \bar{\nu}_\mu$	494
$\frac{1}{2}$	$\frac{1}{2}$	-1	\bar{K}^0	$ \bar{d}s\rangle$	$\bar{K}^0 \rightarrow \pi\pi(\pi)$	498
0	0	0	η	$\sqrt{\frac{1}{6}}(d\bar{d}\rangle + u\bar{u}\rangle - 2 s\bar{s}\rangle)$	$\eta \rightarrow \gamma\gamma$	549
0	0	0	η'	$\sqrt{\frac{1}{3}}(d\bar{d}\rangle + u\bar{u}\rangle + s\bar{s}\rangle)$	$\eta' \rightarrow \eta\pi\pi$ or $\rho\gamma$	958

Table 5.1 Properties of the SU(3) pseudoscalar meson nonet states.

I	I_3	S	Meson	Composition	Decay	Mass (MeV)
1	1	0	ρ^+	$ u\bar{d}\rangle$		
1	-1	0	ρ^-	$ \bar{u}d\rangle$	$\rho \rightarrow \pi\pi$	776
1	0	0	ρ^0	$\sqrt{\frac{1}{2}}(d\bar{d}\rangle - u\bar{u}\rangle)$		
$\frac{1}{2}$	$\frac{1}{2}$	1	K^{*+}	$ u\bar{s}\rangle$		
$\frac{1}{2}$	$-\frac{1}{2}$	1	K^{*0}	$ d\bar{s}\rangle$		
$\frac{1}{2}$	$-\frac{1}{2}$	-1	K^{*-}	$ \bar{u}s\rangle$	$K^* \rightarrow K\pi$	892
$\frac{1}{2}$	$\frac{1}{2}$	-1	\bar{K}^{*0}	$ \bar{d}s\rangle$		
0	0	0	ω	$\sqrt{\frac{1}{2}}(d\bar{d}\rangle + u\bar{u}\rangle)$	$\omega \rightarrow \pi^+\pi^-\pi^0$	783
0	0	0	ϕ	$ s\bar{s}\rangle$	$\phi \rightarrow KK$	1019

Table 5.2 Properties of the SU(3) vector meson nonet states.

such that the quark composition is given by

$$\begin{aligned}
 |\rho^0\rangle &= \sqrt{\frac{1}{2}}(|d\bar{d}\rangle - |u\bar{u}\rangle), & \text{BR}(\rho^0 \rightarrow \pi^+\pi^-) &= 100\% \\
 |\omega\rangle &\approx \sqrt{\frac{1}{2}}(|d\bar{d}\rangle + |u\bar{u}\rangle), & \text{BR}(\omega \rightarrow \pi^+\pi^-\pi^0) &= 90\% \\
 |\phi\rangle &\approx |s\bar{s}\rangle, & \text{BR}(\phi \rightarrow K\bar{K}) &= 84\% \\
 & & \text{BR}(\phi \rightarrow \pi^+\pi^-\pi^0) &= 15\%
 \end{aligned}$$

5.2.4 Baryons

The baryon wavefunction is made of four parts,

$$\Psi = \psi_{\text{space}}\psi_{\text{spin}}\psi_{\text{flavour}}\psi_{\text{colour}}$$

and we consider only ground-state baryons (p , n , Δ , etc.) that have no orbital angular momentum ($L = 0$). ψ_{space} is symmetric. The colour wavefunction ψ_{colour} is always antisymmetric:

$$\psi_{\text{colour}} = \sqrt{\frac{1}{6}}(|RGB\rangle + |GBR\rangle + |BRG\rangle - |GRB\rangle - |BGR\rangle - |RBG\rangle)$$

With $\psi_{\text{space}}\psi_{\text{colour}}$ being antisymmetric and the overall fermionic wavefunction required to be antisymmetric, $\psi_{\text{spin}}\psi_{\text{flavour}}$ must be symmetric.

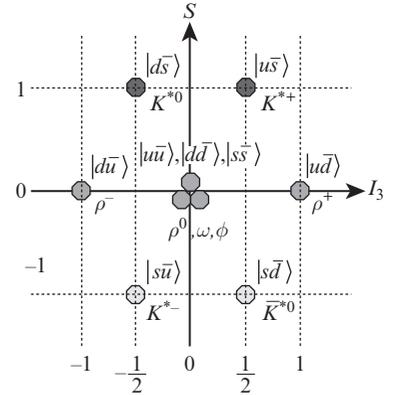


Fig. 5.5 SU(3) vector meson nonet states as functions of I_3 and S .

ψ_{spin} and ψ_{flavour} may be a mixture of symmetric and antisymmetric states, but the total spin–flavour wavefunction must be symmetric.

- We deal with ψ_{spin} first. The possible combinations are

$$2 \otimes (2 \otimes 2) = 8 = \underbrace{4}_{\text{sym}} \oplus \underbrace{2}_{\text{mix}_S} \oplus \underbrace{2}_{\text{mix}_A}$$

where the four symmetric states (‘sym’) have spin $\frac{3}{2}$ and the other two (‘mix_S’ and ‘mix_A’), with mixed symmetry, have spin $\frac{1}{2}$.

- The SU(3) decomposition of three flavours is found by first combining two quark states, then adding the third:

$$3 \otimes 3 = \underbrace{6}_{\text{sym}} \oplus \underbrace{3}_{\text{asym}}$$

$$3 \otimes 3 \otimes 3 = \underbrace{10}_{\text{sym}} \oplus \underbrace{8}_{\text{mix}_S} \oplus \underbrace{8}_{\text{mix}_A} \oplus \underbrace{1}_{\text{asym}}$$

where ‘sym’, ‘mix_S’, and ‘mix_A’ are as defined above and ‘asym’ is the totally antisymmetric three-quark combination.

- Finally, the possible symmetric (SU(3), SU(2)) combinations are the symmetric (10, 4) and the mixed symmetric–antisymmetric $\sqrt{\frac{1}{2}}[(8, 2)_{\text{mix}_S} + (8, 2)_{\text{mix}_A}]$.

The ground-state octet (spin $\frac{1}{2}$) and decuplet (spin $\frac{3}{2}$) are shown in Fig. 5.6.

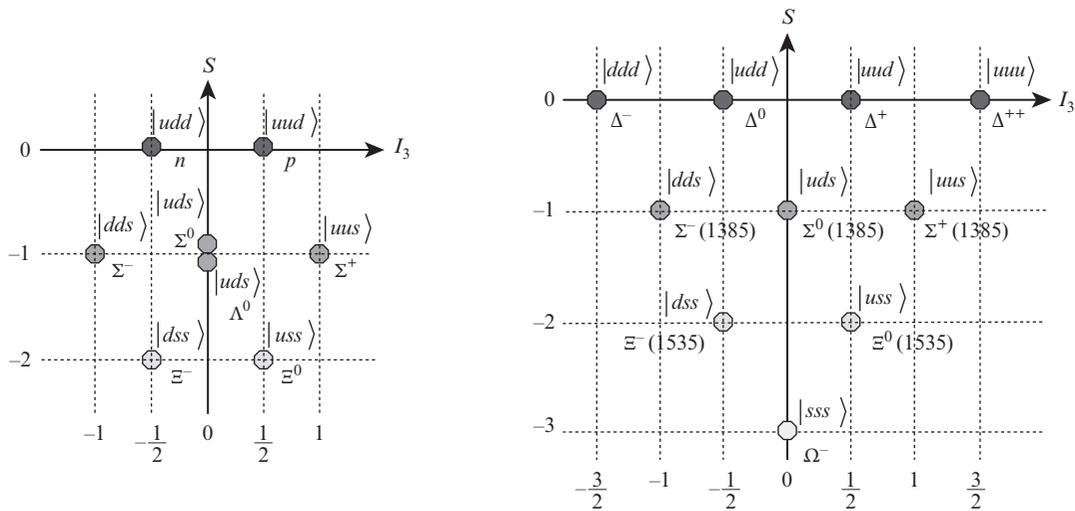


Fig. 5.6 SU(3) ground state (spin- $\frac{1}{2}$) baryon octet states and the (spin- $\frac{3}{2}$) decuplet first excited states.

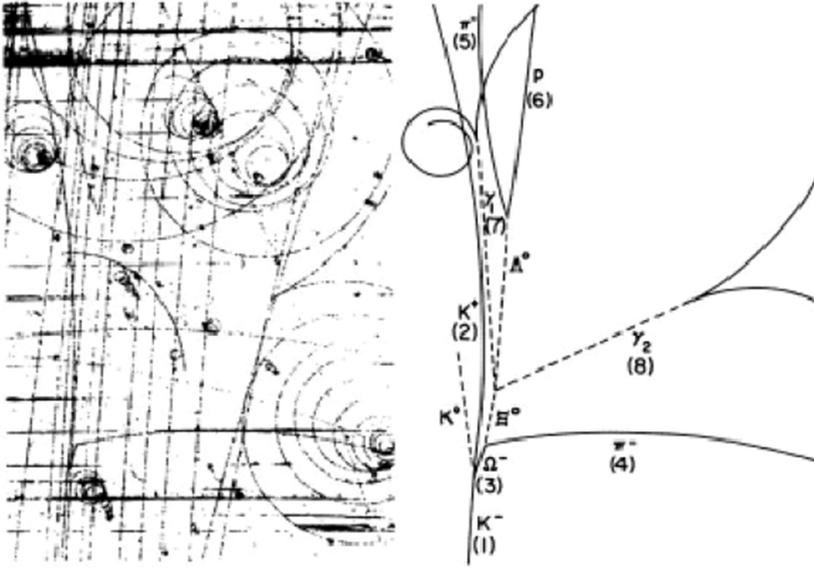
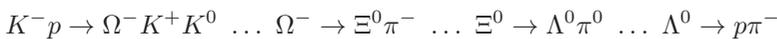


Fig. 5.7 Photograph and line diagram of an Ω^- event from [44].

There are two final points:

- The Δ^{++} is manifestly symmetric in $\psi_{\text{flavour}}\psi_{\text{spin}}\psi_{\text{space}}$. This observation was the driving force for an additional degree of freedom—the *colour charge*.
- The quark model was used to predict the Ω^- , a strangeness-(-3) baryon.¹²

The Ω^- was discovered using K^-p interactions. A bubble chamber photograph [44] of the first observation of an Ω^- is shown in Fig. 5.7. The new particle was observed to decay via three weak decays. The decays are clearly weak because the intermediate particles travel an appreciable distance before decaying in turn. This implies a longer lifetime than decays via the electromagnetic or strong interactions. The production and decay chain is



where each decay involves a decrease in strangeness.¹³

¹²To be historically accurate, (broken) SU(3) flavour symmetry was used. This included quark-like objects, but did not require that they were real particles. Remember, this was 10 years before the key papers on QCD and confinement were published.

¹³ Ω^- has strangeness -3 and spin $\frac{3}{2}$.

5.2.5 Deriving the complete spin-flavour wavefunction

A final step is to obtain the explicit quark model wavefunction for the spin-up proton. We concentrate on the non-trivial $\psi_{\text{spin}}\psi_{\text{flavour}}$ parts, which must be symmetric. From an inspection of the three spin- $\frac{1}{2}$ combinations, for angular-momentum spin (Section 5.1.2) and isospin (end of Section 5.2.1), we note that the $|S = \frac{1}{2}, S_z = \frac{1}{2}\rangle$, and $|I = \frac{1}{2}, I_3 = \frac{1}{2}\rangle$

parts are of mixed symmetry. So we must take care to combine these wavefunctions appropriately. Using \mathcal{M}_S and \mathcal{M}_A as generic labels for symmetric and antisymmetric combinations, respectively, we get

$$|p^\uparrow\rangle = \sqrt{\frac{1}{2}} [\psi_{\text{spin}}(\mathcal{M}_S)\psi_{\text{flavour}}(\mathcal{M}_S) + \psi_{\text{spin}}(\mathcal{M}_A)\psi_{\text{flavour}}(\mathcal{M}_A)]$$

where

$$\psi_{\text{spin}}(\mathcal{M}_S) = \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{6}} (2|\uparrow\uparrow\downarrow\rangle - |\downarrow\uparrow\uparrow\rangle - |\uparrow\downarrow\uparrow\rangle)$$

$$\psi_{\text{spin}}(\mathcal{M}_A) = \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} (|\uparrow\downarrow\uparrow\rangle - |\downarrow\uparrow\uparrow\rangle)$$

and, with different notation, but identical in content:

$$\psi_{\text{flavour}}(\mathcal{M}_S) = \sqrt{\frac{1}{6}} (2|uud\rangle - |duu\rangle - |udu\rangle)$$

$$\psi_{\text{flavour}}(\mathcal{M}_A) = \sqrt{\frac{1}{2}} (|udu\rangle - |duu\rangle).$$

Putting this all together, we can write out the complete proton wavefunction:

$$\begin{aligned} |p^\uparrow\rangle = & \sqrt{\frac{1}{18}} (2|u^\uparrow u^\uparrow d^\downarrow\rangle - |u^\uparrow u^\downarrow d^\uparrow\rangle - |u^\downarrow u^\uparrow d^\uparrow\rangle \\ & + 2|u^\uparrow d^\downarrow u^\uparrow\rangle - |u^\uparrow d^\uparrow u^\downarrow\rangle - |u^\downarrow d^\uparrow u^\uparrow\rangle \\ & + 2|d^\downarrow u^\uparrow u^\uparrow\rangle - |d^\uparrow u^\downarrow u^\uparrow\rangle - |d^\uparrow u^\uparrow u^\downarrow\rangle) \\ & \times |\psi_{\text{colour}}\rangle \end{aligned}$$

5.3 Heavy quarks

By the mid-1960s, broken SU(3) flavour symmetry was becoming established as a plausible explanation for the patterns appearing among the zoo of hadronic states being discovered.¹⁴ The underlying quark model was much more contentious—it was not clear if a quark, which had never been observed in isolation, was a real particle or just a convenient mathematical construct. The experimental and theoretical evidence for quarks being real particles is discussed in Chapter 9. Flavour SU(3), as discussed in this chapter, was used to make successful predictions for hadrons composed of the light u , d , and s quarks. With the discovery of charm and the c quark and some years later of the b quark, a whole new world of charm and beauty hadrons opened up. Although most of the experimental work was carried out at e^+e^- colliders, the first evidence for b quarks came from a fixed-target proton beam on a nuclear target. At the time, it was confidently expected, on the basis that there should be six types of quark to match the six leptons,¹⁵ that the top

¹⁴The zoo continues to grow as more precise experiments coupled with more sophisticated analysis techniques uncover the states with non-zero orbital angular momentum among the quark constituents.

¹⁵The tau lepton with a presumed tau neutrino were discovered in 1975 by the SLAC/Berkeley collaboration using the e^+e^- collider (SPEAR) through a larger rate for $e^+e^- \rightarrow e^\pm\mu^\mp + X^0$ (where X^0 is missing energy) than could be explained by higher-order purely electromagnetic processes.

(t) quark would be found at the higher-energy e^+e^- colliders then being constructed in Europe and Japan, but it was not to be. The first evidence for the t quark came from the Tevatron using $p\bar{p}$ collisions at 1.8 TeV centre-of-mass energy. This section covers the discovery and properties of the b and c quarks; the discovery of the t quark and its properties are covered in detail in Section 8.6.

5.3.1 The charm quark

The need for a fourth quark was already being discussed (see Chapter 7) before the J/ψ and the D mesons were discovered.¹⁶ Indirect evidence for charm arrived in 1974 from two experiments:

- Richter et al. [40], using e^+e^- collisions at SPEAR (the Stanford Positron Electron Asymmetric Ring) with CMS energy $\sqrt{s} \approx 3$ GeV, measured the three processes $e^+e^- \rightarrow$ hadrons, $e^+e^- \rightarrow \mu^+\mu^-$, and $e^+e^- \rightarrow e^+e^-$ (see Fig. 5.8(a)).
- Ting et al. [39] used 28 GeV protons on a beryllium target at the Brookhaven National Laboratory to study the invariant mass spectrum of e^+e^- pairs produced by $p + \text{Be} \rightarrow e^+e^- + X$ (see Fig. 5.8(b)).

¹⁶The J/ψ is a $c\bar{c}$ state and thus has net charm zero, D mesons are $c\bar{q}$ or $q\bar{c}$ with charm quantum number ± 1 .

Ting et al. named the new state J ; Richter et al. chose ψ (a rare example of a particle looking like its name—Fig. 5.8(c) is an event display from the Stanford group [9] showing $\psi' \rightarrow \psi\mu^+\mu^-$ followed immediately by $\psi \rightarrow \mu^+\mu^-$). What was striking in both cases was how exceptionally narrow the Breit–Wigner resonance was to fit the data in the vicinity of

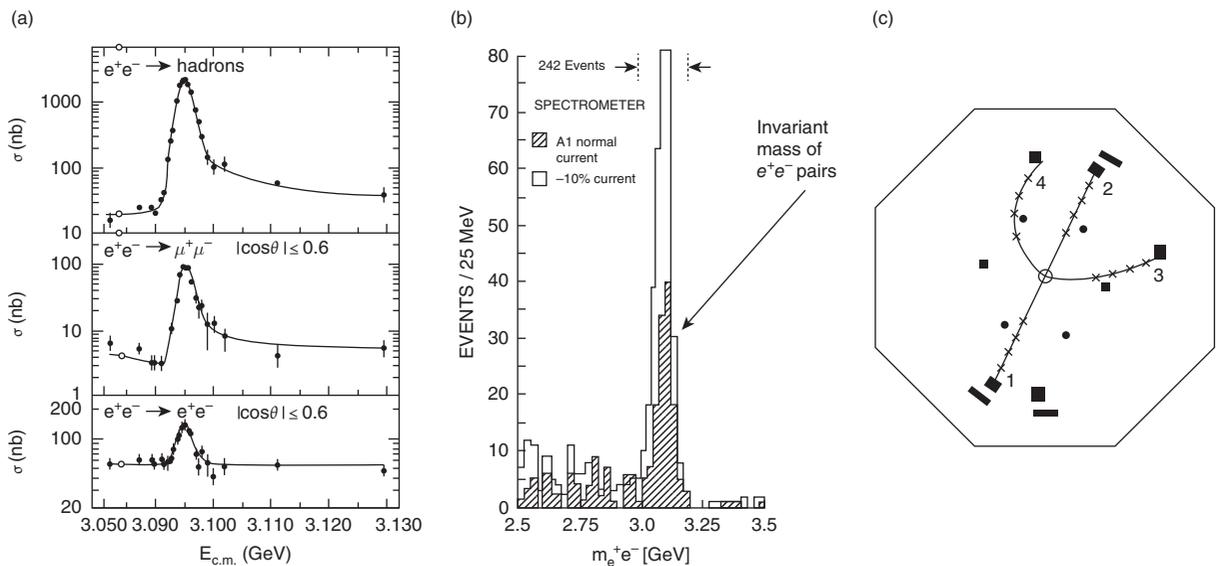


Fig. 5.8 Plots and pictures from the J/ψ discovery papers—a striking object indeed! (a) from [40], (b) from [39], and (c) from [9].

3 GeV. The J/ψ mass and width are 3096 MeV and 93 keV; compare this with a typical hadronic resonance, the $\rho(1700)$, with mass ~ 1720 MeV and width ~ 250 MeV. The e^+e^- and $\mu^+\mu^-$ decay modes of the J/ψ were equal and together amounted to $\sim 12\%$, with hadronic modes accounting for the other 88% of decays. There was no doubt that the J/ψ was a hadronic state, but its decays were highly suppressed—by a factor of roughly 1/2500! Many models were suggested—from new quarks to supersymmetry, but the simplest turned to be the former and it was named the charm quark, with charge $+\frac{2}{3}$ and carrying the new charm quantum number in analogy to strangeness.¹⁷ Data taken at e^+e^- colliders at energies above the J/ψ (and $\psi'(3686)$) showed evidence for a threshold being passed, with new meson states being pair-produced, consistent with $c\bar{q} + \bar{c}q$ pairs.

¹⁷The correct PDG names for the J/ψ and ψ' are $J/\psi(1S)$ and $\psi(2S)$, respectively. The open-charm mesons are the D and \bar{D} states.

5.3.2 The beauty quark

A Fermilab experiment [86] using the 400 GeV proton beam on a nuclear target ($A = \text{Cu}$ or Pt) measured the dimuon invariant mass spectrum in $p + A \rightarrow \mu^+\mu^- + X$. It was expected and observed that the dimuon mass spectrum would fall exponentially. The rapidity-invariant mass double differential cross section was fitted at $y = 0$ by an expression of the form

$$\left. \frac{d^2\sigma}{dm dy} \right|_{y=0} = Ae^{-bm}$$

where m is the invariant mass. A fit to the mass range $6 \text{ GeV} < m < 12 \text{ GeV}$, excluding the range 8.8–10.6 GeV, gave $b = 0.98 \pm 0.02 \text{ GeV}^{-1}$. A statistically significant enhancement was observed at ~ 9.5 GeV. The experiment did not have sufficient mass resolution to resolve the excess above the steeply falling dimuon mass, but it could be fitted with one or two resonances.

The beauty quark is also known as the bottom quark, since it forms a doublet with the top quark.

5.3.3 The top quark

The top quark was discovered at the Tevatron in $p\bar{p} \rightarrow t\bar{t} + X$ at $\sqrt{s} = 1.8 \text{ TeV}$. Its discovery and properties are covered in detail in Section 8.6.

5.3.4 Charm and beauty states

Both the c and b quarks can form meson and baryon states with the light quarks. The t quark decays too rapidly to form such bound states.¹⁸ This section gives a brief overview of heavy flavour states. The charm and beauty states, particularly the B mesons, have been studied in great

¹⁸See Exercise 5.5 at the end of the chapter.

detail at dedicated e^+e^- colliders, providing a wealth of experimental results, only a fraction of which can be covered here. Oscillation phenomena and mixing of $D^0-\bar{D}^0$ and $B^0-\bar{B}^0$ states are discussed in some detail in Chapter 10.

Meson states

The heavy meson states are $Q\bar{q}$ systems; the lightest with zero orbital angular momentum have $J^P = 0^-, 1^-$. Table 5.3 shows those for $c\bar{q}$ and the antiparticles $\bar{c}q$.

For the D mesons,

$$m_{D^\pm} - m_{D^0} = 4.77 \pm 0.10 \text{ MeV}$$

and

$$|m_{D_1^0} - m_{D_2^0}| = 2.39_{-0.63}^{+0.59} \hbar \text{ s}^{-1}$$

Similar details are given for the $b\bar{q}$ and $\bar{b}q$ mesons in Table 5.4. For the B mesons,

$$m_{B^0} - m_{B^\pm} = 0.33 \pm 0.06 \text{ MeV}$$

and

$$\begin{aligned} |m_{B_H^0} - m_{B_L^0}| &= (0.507 \pm 0.005) \times 10^{-12} \hbar \text{ s}^{-1}, \quad \text{or, equivalently,} \\ &= (3.337 \pm 0.033) \times 10^{-10} \text{ MeV} \end{aligned}$$

State		J^P	I	Mass (MeV)	Lifetime ($\times 10^{-15}$ s) or Width
D^+	$c\bar{d}$	0^-	$\frac{1}{2}$	1869.60 ± 0.16	1040 ± 7
D^-	$\bar{c}d$	0^-	$\frac{1}{2}$	1869.60 ± 0.16	1040 ± 7
D^0	$c\bar{u}$	0^-	$\frac{1}{2}$	1864.83 ± 0.14	410 ± 1.5
\bar{D}^0	$\bar{c}u$	0^-	$\frac{1}{2}$	1864.83 ± 0.14	410 ± 1.5
$D^*(2007)^0$	$c\bar{u}$	1^-	1	2006.96 ± 0.16	$\Gamma < 2.1 \text{ MeV}$
$D^*(2010)^\pm$	$c\bar{d}$	1^-	1	2010.25 ± 0.14	$\Gamma = 96 \pm 22 \text{ keV}$

Table 5.3 Lowest-lying charmed meson states; the c quark has charm $C = +1$.

State		J^P	I	Mass (MeV)	Lifetime ($\times 10^{-12}$ s)
B^\pm	$u\bar{b}, \bar{u}b$	0^-	$\frac{1}{2}$	5279.17 ± 0.29	1.638 ± 0.011
B^0, \bar{B}^0	$d\bar{b}, \bar{d}b$	0^-	$\frac{1}{2}$	5279.50 ± 0.30	1.525 ± 0.009

Table 5.4 Lowest-lying beauty meson states; the b quark has beauty $B = +1$.

As for the neutral K mesons, the question arises of CP violation and flavour oscillations. These in turn depend on the small mass difference between the states equivalent to the K_S^0, K_L^0 states. CP violation and charm or beauty oscillations for neutral meson states depend very sensitively on the mass differences; these questions are covered in Chapter 10.

5.3.5 Heavy $Q\bar{Q}$ systems

Positronium—the electromagnetic bound system of an electron and a positron—has provided a very ‘clean’ laboratory for understanding quantum electrodynamics (QED) without any nuclear complications. Similarly, the $Q\bar{Q}$ systems provide a laboratory for understanding QCD. The mass scale provided by the heavy quarks means that perturbative methods can be used for the QCD calculations. The success of these calculations in describing the details of the $Q\bar{Q}$ systems was important for the development of QCD itself. Of the three heavy quarks, only $c\bar{c}$ and $b\bar{b}$ systems exist. The $t\bar{t}$ bound system does not have time to form before the top quarks have decayed.

5.3.6 Charmonium

The first plots with evidence for the J/ψ are shown in Fig. 5.8. Once the excitement of the discovery in quick succession of the J/ψ and ψ' died down, the focus turned to establishing the properties of these states.

Isospin

The first question to be asked is how much u -ness and d -ness there are in this new state. The J/ψ isospin assignment comes from observing

$$\text{BR}(J/\psi \rightarrow \rho^+ \pi^-) = \text{BR}(J/\psi \rightarrow \rho^0 \pi^0) = \text{BR}(J/\psi \rightarrow \rho^- \pi^+)$$

Consulting the Clebsch–Gordan tables gives the J/ψ isospin: $|I = 0, I_3 = 0\rangle$ (i.e. zero isospin!):

	m_1	m_2	$J:$	$M:$	
$ \rho^+\pi^-\rangle = 1, 1\rangle 1, -1\rangle$	1	-1	2	0	$\sqrt{\frac{1}{6}} \quad \sqrt{\frac{1}{2}} \quad \sqrt{\frac{1}{3}}$
$ \rho^0\pi^0\rangle = 1, 0\rangle 1, 0\rangle$	0	0	1	0	$\sqrt{\frac{2}{3}} \quad 0 \quad -\sqrt{\frac{1}{3}}$
$ \rho^-\pi^+\rangle = 1, -1\rangle 1, 1\rangle$	-1	1	0	0	$\sqrt{\frac{1}{6}} \quad -\sqrt{\frac{1}{2}} \quad \sqrt{\frac{1}{3}}$

J^{PC}

Using SPEAR, the SLAC/LBL team were examining $e^+e^- \rightarrow \mu^+\mu^-$. There are two possible processes:

- (1) resonant via a J/ψ decaying electromagnetically;
- (2) non-resonant via a virtual photon.

The lower two plots in Fig. 5.8(a) show $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ and $\sigma(e^+e^- \rightarrow e^+e^-)$ at CMS energies around 3.095 GeV. The middle plot shows a clear dip on the low-mass side of the resonance. This is evidence of the interference between the resonant (via the J/ψ) and non-resonant (via a photon) channels, which can only happen if the J/ψ has the same quantum number as the photon: $J^{PC} = 1^{--}$.

Width

Consider the Breit–Wigner formula for $e^+e^- \rightarrow J/\Psi \rightarrow e^+e^-$:

$$\sigma(E) = 4\pi\lambda^2 \frac{2J+1}{(2s_1+1)(2s_2+1)} \frac{\Gamma_{ee}^2/4}{(E-E_R)^2 + \Gamma^2/4} \quad (5.1)$$

For this reaction, $J = 1, s_1 = s_2 = \frac{1}{2}$, so

$$\sigma(E)_{J/\psi \rightarrow e^+e^-} = 3\pi\lambda^2 \frac{\Gamma_{e^+e^-}^2/4}{(E-E_R)^2 + \Gamma^2/4} \quad (5.2)$$

This can be integrated to find the total cross section, giving

$$\sigma = \int_0^\infty \sigma(E)_{J/\psi \rightarrow e^+e^-} dE = \frac{3\pi^2\lambda^2}{2} \left(\frac{\Gamma_{e^+e^-}}{\Gamma} \right)^2 \Gamma \quad (5.3)$$

From the measurements made by Richter et al. (shown in Fig. 5.8(a)), we can deduce the following:

- The integrated cross section for $e^+e^- \rightarrow J/\psi \rightarrow e^+e^- \approx 800$ nb MeV.
- $\Gamma_{e^+e^-}/\Gamma = \text{BR}(J/\psi \rightarrow e^+e^-) \approx 6\%$.
- Taken together, this gives a value of $\Gamma \approx 93$ keV.

This is much narrower than the full widths of the established $J^{PC} = 1^{--}$ mesons: $K^*(1410)$ with width 232 MeV, $\rho(770)$, with width 149 MeV, and even $\phi(1020)$ with the relatively narrow width of 4.3 MeV. This shows that the J/Ψ cannot be composed of combinations of the light (u, d, s) quarks. The very narrow width can be understood if the state is made of heavier quarks (charm).

Charmonium states

Soon after the discovery of the J/ψ , more $J^{PC} = 1^{--} c\bar{c}$ states were found:

$$\begin{aligned} \psi(2S), \quad \Gamma = 320 \text{ keV}, \quad & \text{BR}(\psi(3686) \rightarrow J/\psi\pi\pi) \approx 50\% \\ \psi(3S), \quad \Gamma = 27.3 \text{ MeV}, \quad & \text{BR}(\psi(3770) \rightarrow D\bar{D}) \approx 85\% \end{aligned}$$

Further $c\bar{c}$ states with other values of J^{PC} were later discovered: the charmonium spectrum is shown in Fig. 5.9.

Charmonium decays are governed by kinematics:

- $m(J/\psi) < m(\psi') < 2m(D) < m(\psi'')$.
- An odd number of gluons is required, Two or three gluons can be in a colour singlet state, so that a decay to two or three gluons would be compatible with colour conservation. For the analogous decays to photons, we can see that C-parity requires three photons. Gluons are coloured and are therefore not eigenstates of C-parity, however. It turns out the two-gluon decay mode is also forbidden, like the photon.
- A single gluon is not possible, since the final state must be colourless.
- Therefore, the minimum number of gluons for a strong decay is three (Fig. 5.10(a)).
- This means that the hadronic decay rate depends on α_s^6 (at the J/ψ mass scale), where α_s is the squared strong coupling constant, in analogy to the fine-structure constant α_{QED} .

The rapid decay of $\psi(3770) \rightarrow D\bar{D}$ is via a single gluon exchange, $\propto \alpha_s^2$, since it is above threshold for strong decays to a pair of charmed mesons (Fig. 5.10(b)).

To see that we need 3 gluon decay modes, first consider the decay of the J/Ψ . This is below threshold to decay into charm mesons. There is

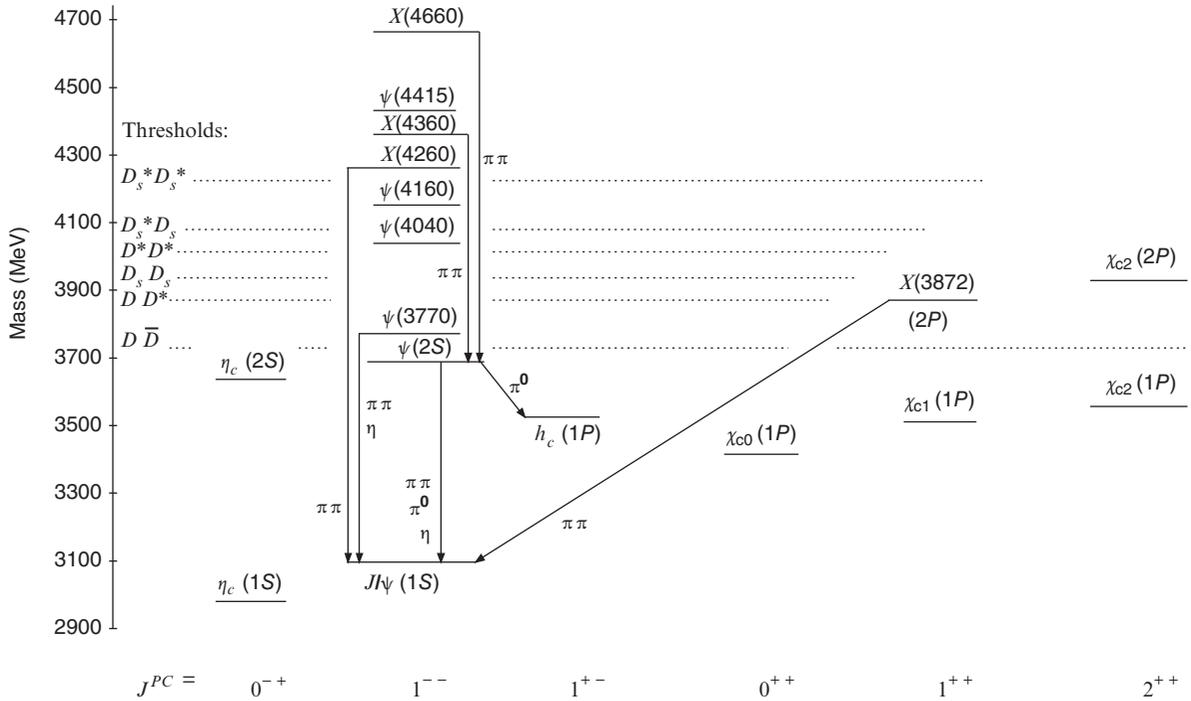


Fig. 5.9 The $c\bar{c}$ charmonium energy levels. From the PDG diagrams [114, p. 1040].

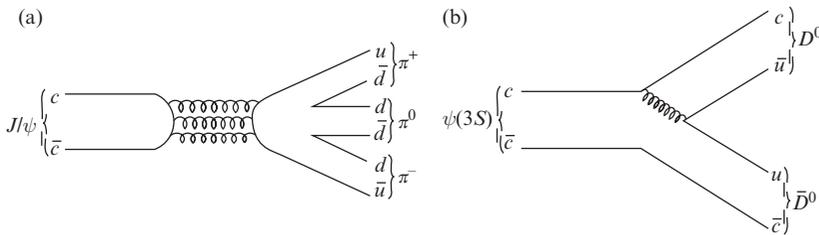


Fig. 5.10 (a) J/ψ decay to a three-pion hadronic final state via three gluons. (b) $\psi(3S)$ decay to $D\bar{D}$.

another hadronic decay, into pions. Because there are no charm quarks in the final state the charm/anticharm pair in the initial state must annihilate for this decay, and there must be gluon propagators connecting the initial and the final state. As this is again a strong decay the conservation laws for the strong interaction must be obeyed at all stages of the process. The relevant conserved property is the charge conjugation eigenvalue. For the initial state with $L = 0$ and $S = 1$, $C = (-1)^{0+1} = -1$. For the final state a decay into a single pion would not satisfy momentum conservation. A decay into $\pi\pi$ ($C = +1$) would violate charge conjugation, so the final state must have at least three pions. More important is the structure of the intermediate gluon state. It cannot be a single gluon, because the initial state is a colour singlet (it's a particle), and a single gluon can never be a colour singlet. The gluon is not an eigenstate to charge conjugation, because of its colour content (e.g. a gluon

with $r\bar{b}$ becomes $-b\bar{r}$, the $-$ sign occurs for similar reasons as the $-$ sign in the charge conjugation eigenvalue for the photon), but a multi-gluon state can be a charge conjugation eigenstate. A two-gluon colour singlet state will contain contributions like $(r\bar{b})(b\bar{r})$. The charge conjugate state will then have $(-b\bar{r})(-r\bar{b})$, so the charge conjugation eigenvalue of the two-gluon state will be $C = +1$, and there can be no strong decay of a J/Ψ into two gluons. A three-gluon state would contain elements like $(r\bar{b})(b\bar{g})(g\bar{r}) \pm (b\bar{r})(g\bar{b})(r\bar{g})$. One of these combinations will have $C = -1$, and so a decay into three gluons is possible.

5.3.7 Comparison with positronium

Positronium

- Positronium is an e^+e^- bound system analogous to the hydrogen atom.
- The energy levels are predicted by the non-relativistic Schrödinger equation with the Coulomb potential, but with a reduced mass $m_e/2$.
- Singlet 1S_0 and triplet 3S_1 states are split by spin-spin interactions.
- States with the same principal quantum number are split by spin-orbit interaction.

Differences between charmonium and positronium

- In charmonium, the size of the strong coupling α_s compared with the electromagnetic α_{QED} leads to larger splitting from spin-spin and spin-orbit interactions than would be the case in atomic physics;
- The potential in charmonium is not $-\alpha_{\text{QED}}/r$ but $\alpha_s/r + \kappa r$. At short distances, the potential is Coulomb-like, but at large distances, the linear confining term dominates.

5.3.8 Bottomonium

A comparison of the ratio

$$R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}$$

in the regions near the ρ^0 , J/ψ , and Υ resonances is shown in Fig. 5.11. Following the discovery of the narrow excess in μ -pair invariant mass distribution measured in $p + A \rightarrow \mu^+\mu^- + X$ already described, the region of 9–10 GeV in centre-of-mass energy was studied by the e^+e^- colliders at SLAC, Cornell, and DESY, where the better resolution of these colliders¹⁹ enabled the system to be resolved into three narrow resonances $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$, with masses of 9.46, 10.02, and 10.36 GeV, respectively. A broader resonance, the $\Upsilon(4S)$ with

¹⁹In an e^+e^- collider, the energy of a resonance can be determined by the beam energy, which can be measured very precisely. In a hadron production experiment, the resonance energy is determined by the final-state particles (electrons or muons) and the resolution is lower.

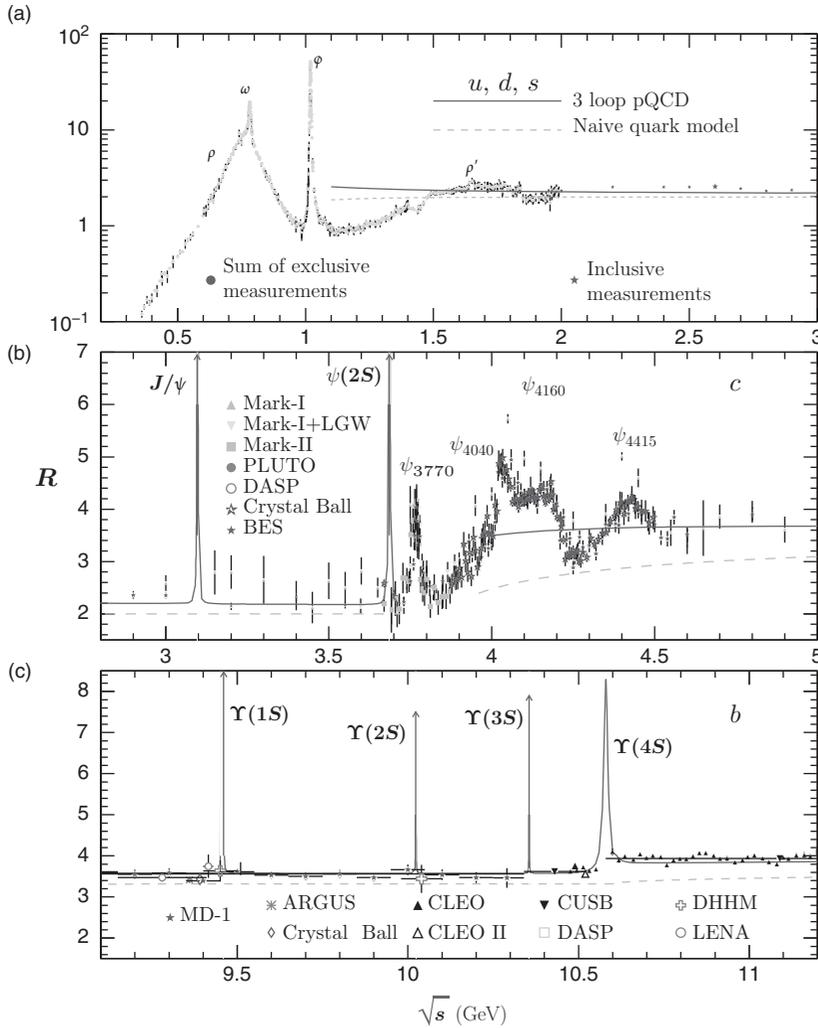


Fig. 5.11 The ratio R in the e^+e^- CMS energy regions: (a) near the ρ ; (b) near the J/ψ ; (c) near the Υ . From the PDG data plots [115, Fig. 50.6].

mass 10.58 GeV, decayed to mesons via $B\bar{B}$ pairs, each with a mass ~ 5.28 GeV. The bottomonium spectrum is shown in Fig. 5.12.

5.4 Exotic hadrons

The Standard Model allows more possibilities for hadrons than considered so far. We need to add gluons to quarks as building blocks, considering different ways in which gluon fields can be configured ‘connecting’ or ‘gluing’ quarks while making sure that the outcome is colourless.

The simplest object is called a glueball, a quarkless set of gluons that is colourless as a whole. No glueball has been unambiguously identified so far; expected lifetimes are short and glueballs would couple easily to conventional mesons, making unambiguous identification difficult.

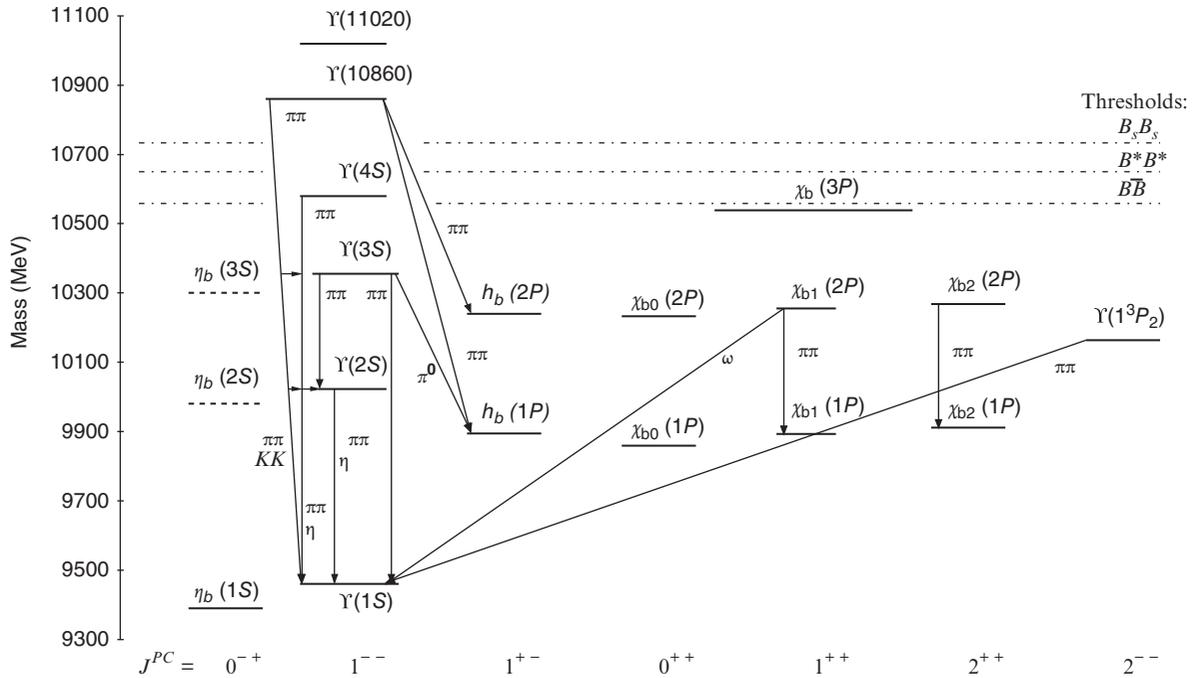


Fig. 5.12 The $b\bar{b}$ bottomonium states. From the PDG diagrams [114, p. 1109].

Having glue fields inside a hadron leads to hybrids. It is possible to excite gluonic degrees of freedom, making gluonic fields vibrate for example. No hybrids have been found so far.

Then there are tetraquarks, four-quark systems such as $qq\bar{q}\bar{q}$. They could be of two types: either a ‘molecule’ of two mesons, with one $q\bar{q}$ orbiting another $q\bar{q}$, or a diquark system, with a qq diquark binding to a $\bar{q}\bar{q}$ antiquark (known also as baryonium). There are some candidates for tetraquark states. The most promising one, not matching the mass, lifetime, and other quantum numbers of any conventional meson (known or predicted), is the narrow $X(3872)$ state discovered by the Belle experiment. The $X(3872)$ can decay to $\pi^+\pi^-J/\psi$ and to $\gamma J/\psi$, suggesting that it contains charm and anticharm quarks. Its quantum numbers, $J^{PC} = 1^{++}$, are now well established by the LHCb experiment. Whether it is a ‘molecule’ or a diquark system or a mixture of states is not yet known.

What tetraquarks are in relation to mesons, pentaquarks are in relation to baryons: $qqqq\bar{q}$. Two states discovered by LHCb, $P_c(4450)^+$ and $P_c(4380)^+$ [106], are good candidates for pentaquarks consisting of $uudc\bar{c}$ quarks. How the quarks are bound remains to be established.

There might be even more complicated systems. For example an equal mixture of u , d , and s quarks could exist as a state as simple as one in which two Λ^0 baryons are bound together (if the mass of such a state were below the Λ^0 -nucleon threshold then its lifetime would be of

the order of days or months), or even macroscopic systems with larger number of quarks. There is also a possibility of an analogue to a neutron star—a quark star.

Chapter summary

- Hadrons as qqq and $q\bar{q}$ states.
- Isospin and SU(3) flavour.
- Flavour quantum numbers: strangeness, charm, beauty.
- Isospin and SU(3) flavour.
- Patterns of meson and baryon states.
- Heavy $Q\bar{Q}$ states: charmonium and bottomonium.
- Exotic hadrons.

Further reading

- Griffiths, D. (2008). *Introduction to Elementary Particles* (2nd edn). Wiley. Chapter 5.
- Perkins, D. H. (1987). *Introduction to High Energy Physics* (3rd edn). Addison-Wesley. Chapter 5.
- Halzen, F. and Martin, A. D. (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley. Chapter 2.
- Martin, B. R. and Shaw, G. (2008). *Particle Physics* (3rd edn). Wiley. Chapter 6.
- Close, F. E. (1979). *An Introduction to Quarks and Partons*. Academic Press. Chapters 1–4 give a more advanced introduction to the quark model and the associated group theory.

Exercises

- (5.1) Using the information in Section 2.2 on addition of angular momentum and Clebsch–Gordan coefficients, check that you can account for all states in the baryon multiplets constructed out of u and d quarks only.
- (5.2) Follow through the calculation of the Δ spin- $\frac{3}{2}$, I-spin- $\frac{3}{2}$ resonance branching ratios to pion–nucleon states.
- (5.3) Referring to Fig. 5.1 and using the PDG data tables, give an explanation of the difference in π^+p and π^-p total cross sections at π^\pm beam energies of around 1 GeV in the fixed-target laboratory frame.
- (5.4) J/ψ , verify the formula for the total cross-section integral given in eqn 5.3. Use it and the data given to estimate the total width of the resonance. Why was it so surprising?
- (5.5) The top quark has a mass of $172.0 \pm 0.9 \pm 1.3$ GeV with an upper limit on its full width of $\Gamma < 13.1$ GeV. Estimate its mean lifetime and, by considering the available phase space for its decays, explain why there are no hadronic states including the top quark.

6

Relativistic quantum mechanics

6.1 Special relativity	140
6.2 One-particle states	147
6.3 The Klein–Gordon equation	149
6.4 The Dirac equation	155
6.5 Gauge symmetry	172
Chapter summary	178
Further reading	178
Exercises	179

The aim of this chapter is to introduce a relativistic formalism that can be used to describe particles and their interactions. The emphasis is on those elements of the formalism that can be carried on to relativistic quantum field theory (RQF), which underpins the theoretical framework of high-energy particle physics.

We begin with a brief summary of special relativity, concentrating on 4-vectors and spinors. One-particle states and their Lorentz transformations follow, leading to the Klein–Gordon and Dirac equations for probability amplitudes, i.e. relativistic quantum mechanics (RQM). Readers who want to get to RQM quickly, without studying its foundation in special relativity, can skip the first sections and start reading from Section 6.3.

Intrinsic problems of RQM are discussed and a region of applicability of RQM is defined. Free-particle wavefunctions are constructed and particle interactions are described using their probability currents. Gauge symmetry is introduced, which allows the interaction between a particle and a classical gauge field to be described within the formalism.

6.1 Special relativity

¹Albert Einstein, 1879–1955.

²Note that in that chapter, transformations are in an ‘active’ sense (i.e. the coordinate system does not change but vectors do) rather than the ‘passive’ sense (i.e. the coordinate system changes but vectors do not) considered in this book. It should also be noted that the metric in [110] is $-+++$, in contrast to the one used in this book, which is $+---$ (see Note 5).

³Greek indices $\mu, \nu = 0, 1, 2, 3$ and Latin indices $i, j = 1, 2, 3$.

Einstein’s¹ special relativity is a necessary and fundamental part of any formalism of particle physics. We begin with a brief summary. For a full account, refer to specialized books, for example [128] or [127]. Theory-oriented students with a good mathematical background might want to consult books on groups and their representations, for example [46], followed by introductory books on RQM/RQF, for example [107]. Here we are only going to present conclusions without derivations, avoiding group-theoretical language and aiming at a presentation of key concepts at a qualitative level. Chapter 41 in [110] on spinors is recommended.²

The basic elements of special relativity are 4-vectors (or, strictly speaking, contravariant 4-vectors) such as a 4-displacement³ $x^\mu = (t, \mathbf{x}) = (x^0, x^1, x^2, x^3) = (x^0, x^i)$ or a 4-momentum $p^\mu = (E, \mathbf{p}) = (p^0, p^1, p^2, p^3) = (p^0, p^i)$. 4-vectors have real components and form a vector space. There is a metric tensor $g_{\mu\nu} = g^{\mu\nu}$ that is used to form a dual space to the space of 4-vectors. This dual space is a vector

space of linear functionals, known as 1-forms (or covariant 4-vectors), which act on 4-vectors. For every 4-vector x^μ , there is an associated 1-form $x_\mu = g_{\mu\nu}x^\nu$. Such a 1-form is a linear functional that, acting on a 4-vector y^μ , gives a real number $= g_{\mu\nu}x^\nu y^\mu$. This number is called the scalar product⁴ $x \cdot y$ of x^ν and y^μ . The Lorentz transformation between two coordinate systems, Λ^μ_ν , with $x'^\mu = \Lambda^\mu_\nu x^\nu$, leaves the scalar product unchanged which is equivalent to $g_{\rho\sigma} = g_{\mu\nu} \Lambda^\mu_\rho \Lambda^\nu_\sigma$.

In the standard configuration, the Lorentz transformation becomes the Lorentz boost along the first space coordinate direction and is given by⁵

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with

$$\beta = \frac{v}{c}, \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

where v is the velocity of the boost.

Two Lorentz boosts along different directions are equivalent to a single boost and a space rotation. This means that Lorentz transformations, which can be seen as space–time rotations, include Lorentz boosts (rotations by a purely imaginary angle) as well as space rotations (by a purely real angle). Representing Lorentz transformations by 4-dimensional real matrices acting on 4-vectors is not well suited for combining Lorentz boosts and space rotations in a transparent way. Even a simple question like ‘What is the single space rotation that is equivalent to a combination of two arbitrary space rotations?’ is hard to answer.

A better way is to represent Lorentz transformations by 2-dimensional complex matrices. First we consider a 3-dimensional real space and rotations. With every rotation in that 3-dimensional real space we can associate a 2×2 complex matrix, called a spin matrix,⁶

$$R = \cos\left(\frac{1}{2}\theta\right) + i \sin\left(\frac{1}{2}\theta\right) (\sigma_x \cos\alpha + \sigma_y \cos\beta + \sigma_z \cos\gamma)$$

or

$$\begin{aligned} R &= \cos\left(\frac{1}{2}\theta\right) + i \sin\left(\frac{1}{2}\theta\right) (\mathbf{n} \cdot \boldsymbol{\sigma}) \\ &= \exp\left[i\left(\frac{1}{2}\theta\right) (\mathbf{n} \cdot \boldsymbol{\sigma})\right] \end{aligned} \quad (6.1)$$

where θ is the angle of rotation, α, β, γ are the angles⁷ between the axis of rotation \mathbf{n} and the coordinate axes, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices. Note that R is unitary: $R^\dagger = R^{-1}$. The vector space of spin matrices (a subspace of all 2×2 complex matrices) is thus defined using four basis vectors, such as the unit matrix and three basis vectors formed

⁴A similar situation occurs in the infinite-dimensional vector space of states in quantum mechanics (with complex numbers there). For every state, represented by a vector known as a ket, for example $|x\rangle$, there is a 1-form known as a bra, $\langle x|$, that, acting on a ket $|y\rangle$, gives a number $\langle x | y \rangle$, which is called the scalar product of the two kets $|x\rangle$ and $|y\rangle$.

⁵The metric is represented by the matrix

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

⁶Also known as Hamilton’s quaternion or a spinor transformation or a rotation operator.

⁷Only two of the angles α, β, γ are independent.

using the Pauli matrices: $i\sigma_x, i\sigma_y, i\sigma_z$. In this basis, the spin matrix R has the coordinates $\cos(\frac{1}{2}\theta)$, $\sin(\frac{1}{2}\theta)\cos\alpha$, $\sin(\frac{1}{2}\theta)\cos\beta$, and $\sin(\frac{1}{2}\theta)\cos\gamma$. Combining two rotations, one multiplies corresponding spin matrices and describes the outcome using the above basis; thus getting all the parameters of the equivalent single rotation.⁸ The next step is to associate each 3-dimensional space (real numbers) vector $\mathbf{x} = (x^1, x^2, x^3)$ with a corresponding spin matrix (there is no unit matrix in the basis here—only the three Pauli matrices):

$$X = x^1\sigma_x + x^2\sigma_y + x^3\sigma_z \quad (6.2)$$

Then, under the space rotation, \mathbf{x} is transformed to \mathbf{x}' and X is transformed to $X' = RXR^\dagger = x'^1\sigma_x + x'^2\sigma_y + x'^3\sigma_z$, from which we can read the coordinates of \mathbf{x}' .

The beauty of this approach is that it extends seamlessly to space–time rotations, i.e. to the Lorentz transformations. The spin matrix R of eqn 6.1 becomes the Lorentz transformation

$$L = \exp\left[(-\boldsymbol{\rho} + i\theta\mathbf{n}) \cdot \frac{1}{2}\boldsymbol{\sigma}\right] \quad (6.3)$$

⁸This representation of rotations is used in video game programming because of this ease of combining rotations quickly.

⁹ \mathbf{n}_ρ is the unit vector along the direction of the Lorentz boost.

where $\boldsymbol{\rho} = \rho\mathbf{n}_\rho$ is the rapidity.⁹ The rapidity is related to the Lorentz β and γ parameters by

$$\tanh\rho = \beta, \quad \cosh\rho = \gamma, \quad \sinh\rho = \beta\gamma$$

Now, a combination of two Lorentz transformations is very transparent—just addition of real and imaginary parts in the exponent. Association of a 4-vector x^μ with a Hermitian spin matrix¹⁰ X ,

$$X = x^0 + x^1\sigma_x + x^2\sigma_y + x^3\sigma_z \quad (6.4)$$

allows us to get its Lorentz-transformed coordinates from $X' = LXL^\dagger = x'^0 + x'^1\sigma_x + x'^2\sigma_y + x'^3\sigma_z$. Finally, the Lorentz boost alone ($\theta = 0$) along \mathbf{n}_ρ is

$$L = \exp\left(-\boldsymbol{\rho} \cdot \frac{1}{2}\boldsymbol{\sigma}\right) = \cosh\left(\frac{1}{2}\rho\right) - \mathbf{n}_\rho \cdot \boldsymbol{\sigma} \sinh\left(\frac{1}{2}\rho\right) \quad (6.5)$$

¹⁰Now with four matrices as the basis and X satisfying $X = X^\dagger$ by definition.

6.1.1 Spinors

¹¹Spinors are vectors in the mathematical sense since they form a complex vector space, but they are not vectors like a displacement \mathbf{x} , because they transform (e.g. under rotation) differently.

Spin matrices can act on 2-component complex vectors called spinors.¹¹ Spinors, like vectors and tensors, are used in a number of different areas of physics, including classical mechanics. They play a particularly important role in RQM and in this section we will describe them in some detail. Under a space rotation R , a spinor ξ (ξ^α to be more precise) transforms in the following way:

$$\xi' = R\xi$$

For comparison, the coordinates of a vector \mathbf{x} transform under a space rotation as

$$X' = R X R^\dagger = x'^1 \sigma_x + x'^2 \sigma_y + x'^3 \sigma_z$$

Thus, in a rotation of the coordinate system by $\theta = 2\pi$, $R = -1$ because the $\frac{1}{2}\theta$ in R gives $\xi' = -\xi$ and $\mathbf{x}' = \mathbf{x}$. Continuing the rotation by a further 2π , so all together by 4π , results in $\xi' = \xi$. Does that counter-intuitive minus sign resulting from the 2π rotation have any physical significance? Yes it does, as was demonstrated in a beautiful experiment [122] using neutrons.

One of two coherent neutron beams passes through a magnetic field of variable strength. In the magnetic field, the neutrons' magnetic moments precess with the Larmor frequency and the angle of the precession is easily calculated as a function of the strength of the magnetic field. After passing through the magnetic field, the beam interferes with the second beam, which followed a path outside the magnetic field. As demonstrated in Fig. 6.1, an angle of 4π is needed for the neutron wavefunction to reproduce itself. A 2π rotation gives a factor -1 in front of the original neutron wavefunction, as predicted for a spin- $\frac{1}{2}$ spinor.

So far, one could think about spinors as being identical with the Pauli spinors¹² of non-relativistic quantum mechanics (NRQM). This is not quite right. The reason is that the Pauli spinors of NRQM live

¹²Eigenstates of spin operators, like the spin projection on the z axis, $\frac{1}{2}\hbar\sigma_z$, for a spin- $\frac{1}{2}$ particle in NRQM [94].

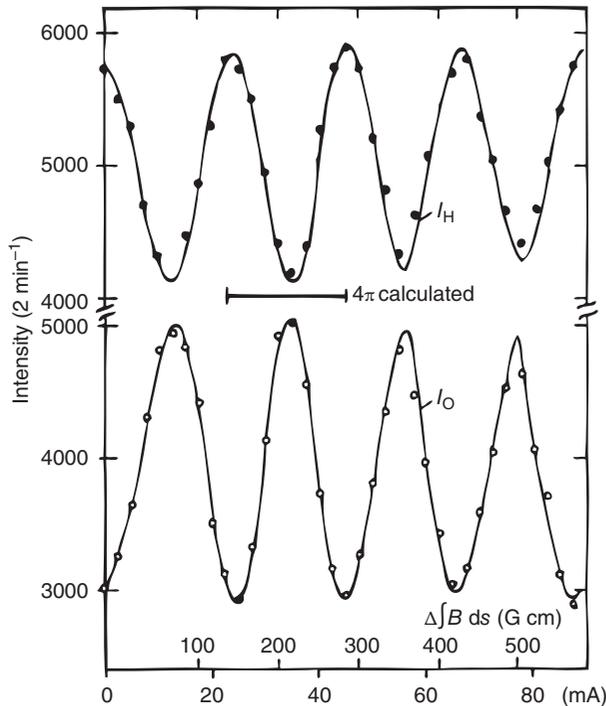


Fig. 6.1 A phase change of 4π is needed to get the same intensity from the interference of two neutron beams. Taken from [122].

¹³Hermann Weyl, 1885–1955.

in space and not in space–time and we do not know how to Lorentz-transform them. What we are constructing now are Weyl¹³ spinors (there is more than one type) living in space–time, and we *do* know how to Lorentz-transform these. Weyl spinors are needed to construct Dirac spinors, or bispinors, since two Weyl spinors of different type are needed for one Dirac spinor.

¹⁴A tensor product or outer product or dyadic product—something like this:

$$\begin{pmatrix} a \\ b \end{pmatrix} (c \ d) = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix}.$$

We will look now at the spin matrices X of eqn 6.4 from a different viewpoint, seeing them as tensors created by a tensor product¹⁴ of 2-dimensional spinors. Weyl spinors are rooted in space–time, not only in space like Pauli spinors. Consider the Lorentz transformation L of a spin matrix X built from spinors $\xi = \begin{pmatrix} a \\ b \end{pmatrix}$ and $\eta = \begin{pmatrix} c \\ d \end{pmatrix}$:

$$X' = \begin{pmatrix} a' \\ b' \end{pmatrix} (c' \ d') = L \begin{pmatrix} a \\ b \end{pmatrix} (c \ d) L^\dagger = LXL^\dagger$$

We can see that $\xi' = L\xi$ but $\eta' = L^*\eta$ (after taking the transpose, $L^{\dagger T} = L^*$). There are two different types of spinors, transforming differently. Those that transform with the complex conjugate L^* are called dotted Weyl spinors, distinguished from the undotted ξ^α by a dot written above the index: $\eta^{\dot{\alpha}}$; for example, $(\xi^\alpha)^*$ is a dotted spinor. The spin matrix X is then written as $X^{\alpha\dot{\beta}}$ ($\alpha = 1, 2$ and $\dot{\beta} = 1, 2$). There is a metric tensor

$$\epsilon_{\alpha\beta} = \epsilon^{\alpha\beta} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

(and an identical one for the dotted spinors) to create a dual space of 1-forms:

$$\xi_\alpha = \epsilon_{\alpha\beta}\xi^\beta \quad (\text{and} \quad \eta_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\eta^{\dot{\beta}})$$

¹⁵Note that this gives $\xi^\alpha\xi_\alpha = 0$ and $\xi^\alpha\zeta_\alpha = -\xi_\alpha\zeta^\alpha$.

(thus $\xi_1 = \xi^2$ and $\xi_2 = -\xi^1$). The scalar product¹⁵

$$\xi^\alpha\zeta_\alpha = \epsilon_{\alpha\beta}\xi^\alpha\zeta^\beta$$

(and similarly for the dotted spinors) is invariant with respect to the Lorentz transformation. Because undotted Weyl spinors and dotted Weyl spinors are different objects, the scalar product, or in general any contraction, can only be performed on the same type of spinors: an undotted index is contracted with another undotted index and a dotted index is contracted with another dotted one; one cannot contract a dotted index with an undotted one.

In order to gain more insight, we go beyond Lorentz transformations and consider space inversion, P : $P(x^0, \mathbf{x}) = (x^0, -\mathbf{x})$. The space inversion P commutes with space rotations, but not with Lorentz transformations, because Lorentz transformations affect the time component and P does not. To illustrate this, consider a boost Λ followed by a space inversion P in 4-dimensional space–time, $P\Lambda$. It is evident that this is equivalent to a space inversion P followed by a boost, $P\Lambda = \Lambda'P$,

but $\Lambda \neq \Lambda'$: if Λ is a boost with velocity \mathbf{v} , then Λ' is a boost with velocity $-\mathbf{v}$. Thus $[P, \Lambda] \neq 0$, and therefore P is not proportional to the identity operator, which commutes with every operator.

We now return to Weyl spinors. Because space inversion is not proportional to the identity operator, space inversion does not transform ξ^α into ξ^α times a number. It transforms ξ^α into a spinor of a different type, which transforms under the Lorentz transformation differently to ξ^α . Just as Pauli spinors represent spin in NRQM, Weyl spinors are going to represent spin in RQM. We know that space inversion leaves spin unaffected, and therefore, under P , ξ^α needs to be transformed to a spinor that transforms under space rotations in the same way as ξ^α and represents the same spin state.¹⁶ Out of all three possibilities, only the 1-form $\eta_{\dot{\alpha}}$ transforms in the same way. So, under space inversion, $\xi^\alpha \rightarrow \eta_{\dot{\alpha}}$ and $\eta_{\dot{\alpha}} \rightarrow \xi^\alpha$. In the discussion on space inversion, $P^2 = 1$ is assumed. This is fine for all particles except Majorana particles. For a Majorana particle, $P^2 = -1$ and the transformation of spinors under P is different to that given here. We will define a Majorana particle later.

As the spinors ξ^α and $\eta_{\dot{\alpha}}$ play a very important role in RQM, the following is a summary of how they behave under various transformations (note here that $(L^\dagger)^{-1} = \epsilon L^* \epsilon^{-1}$):

$$\text{rotation } R : \quad \xi^\alpha \rightarrow R\xi^\alpha, \quad \eta_{\dot{\alpha}} \rightarrow R\eta_{\dot{\alpha}} \quad (6.6)$$

$$\text{Lorentz transformation } L : \quad \xi^\alpha \rightarrow L\xi^\alpha, \quad \eta_{\dot{\alpha}} \rightarrow (L^\dagger)^{-1}\eta_{\dot{\alpha}} \quad (6.7)$$

$$\text{Lorentz boost } L^\dagger = L : \quad \xi^\alpha \rightarrow L\xi^\alpha, \quad \eta_{\dot{\alpha}} \rightarrow L^{-1}\eta_{\dot{\alpha}} \quad (6.8)$$

$$\text{space inversion } P : \quad \xi^\alpha \rightarrow \eta_{\dot{\alpha}}, \quad \eta_{\dot{\alpha}} \rightarrow \xi^\alpha \quad (6.9)$$

Suppose there is a spin- $\frac{1}{2}$ particle with 4-momentum p^μ described in a particular reference frame by $p^{\alpha\dot{\beta}}$ via eqn 6.4.¹⁷ Note that $p^{\alpha\dot{\beta}}$ is identical to $p^{\dot{\beta}\alpha}$; it is not a transpose operation. Following our non-relativistic intuition gained from using Pauli spinors, we want to represent the spin of that particle by ξ^α . In an attempt to write a covariant¹⁸ equation, we could try to contract the undotted index α , but that would lead to something like

$$p_{\alpha\dot{\beta}}\xi^\alpha = m\eta_{\dot{\beta}}$$

where $\eta_{\dot{\beta}}$ is a dotted spinor different from ξ^α related to the uncontracted dotted index, and m is a dimensionful scalar¹⁹ parameter appearing because of the energy dimensionality of p^μ . So the equation is covariant only when $m = 0$, because we do not have any dotted spinor in hand to put on the right-hand side. A similar outcome is obtained if we have only $\eta_{\dot{\beta}}$ instead of ξ^α . Having a column vector with two complex numbers is not enough; we also need to indicate how the two numbers transform under Lorentz transformations. Similarly, for vectors in 3-dimensional space, three real numbers are not enough; we need to know whether they represent a polar or an axial vector, since these transform differently under space inversion. So insisting on only one type of spinor excludes the other type, because they transform differently.

¹⁶For a more complete treatment of the material presented here on spinors, Chapter III of [50] is recommended.

¹⁷We have

$$\begin{aligned} p^{1\dot{1}} &= p_{2\dot{2}} = p^0 + p^3 \\ p^{2\dot{2}} &= p_{1\dot{1}} = p^0 - p^3 \\ p^{1\dot{2}} &= -p_{2\dot{1}} = p^1 - ip^2 \\ p^{2\dot{1}} &= -p_{1\dot{2}} = p^1 + ip^2 \end{aligned}$$

¹⁸Here and in the rest of this chapter, covariant means covariant with respect to Lorentz transformations.

¹⁹Henceforth, scalar means scalar with respect to Lorentz transformations.

Consequently, we end up with two independent Lorentz-invariant Weyl equations:

$$p_{\alpha\dot{\beta}}\xi^\alpha = 0, \quad (p^0 - \mathbf{p} \cdot \boldsymbol{\sigma})\xi = 0 \quad (6.10)$$

$$p^{\alpha\dot{\beta}}\eta_{\dot{\beta}} = 0, \quad (p^0 + \mathbf{p} \cdot \boldsymbol{\sigma})\eta = 0 \quad (6.11)$$

²⁰In quantum mechanics, a helicity operator representing the projection of a particle's spin on the direction of its momentum is defined as

$$\frac{\mathbf{p} \cdot \boldsymbol{\sigma}}{|\mathbf{p}|}$$

For a massless particle, this is equivalent to

$$\frac{\mathbf{p} \cdot \boldsymbol{\sigma}}{p^0}$$

²¹We have $p^{\alpha\dot{\beta}}p_{\gamma\dot{\beta}} = p^\mu p_\mu \delta_\gamma^\alpha$.

In the context of RQM, eqn 6.10 represents an equation of motion for a free massless spin- $\frac{1}{2}$ particle with positive helicity²⁰ and eqn 6.11 an equation of motion for a different free massless spin- $\frac{1}{2}$ particle with negative helicity. Each equation is not covariant under space inversion and violates parity because the space inversion, eqn 6.9, sends each spinor beyond the formalism—only one type of spinor is present in each formalism. At present, there are no known particles that could be described by either of the Weyl equations. If the electron neutrino were exactly massless, it would be described by eqn 6.11 and the hypothetically massless and different electron antineutrino would be described by eqn 6.10.

Suppose now that we have $p^{\alpha\dot{\beta}}$ and two different spinors ξ^α and $\eta_{\dot{\beta}}$ to describe a spin- $\frac{1}{2}$ particle. First, we contract the undotted index, giving $p_{\alpha\dot{\beta}}\xi^\alpha = m\eta_{\dot{\beta}}$, then acting with²¹ $p^{\alpha\dot{\beta}}$ on $\eta_{\dot{\beta}}$ from that equation gives $m\xi^\alpha$ under the condition that $m^2 = p^\mu p_\mu$. The result is a covariant set of equations

$$\begin{aligned} p^{\alpha\dot{\beta}}\eta_{\dot{\beta}} &= m\xi^\alpha, & (p^0 + \mathbf{p} \cdot \boldsymbol{\sigma})\eta &= m\xi \\ p_{\alpha\dot{\beta}}\xi^\alpha &= m\eta_{\dot{\beta}}, & (p^0 - \mathbf{p} \cdot \boldsymbol{\sigma})\xi &= m\eta \end{aligned} \quad (6.12)$$

Requiring that, under space inversion, ξ^α and $\eta_{\dot{\beta}}$ be transformed into each other as in eqn 6.9 makes the set of equations 6.12 invariant under space inversion, because, simultaneously, $p^{\alpha\dot{\beta}}$ and $p_{\alpha\dot{\beta}}$ are also transformed into each other. The spinors ξ^α and $\eta_{\dot{\beta}}$ are combined into a single four-component bispinor called the Dirac²² spinor and the two equations become one equation called the Dirac equation. In the context of RQM, the Dirac equation describes a spin- $\frac{1}{2}$ particle like the electron.

In order to gain more insight into the origin of the Dirac equation, consider the Lorentz boost, eqns 6.5 and 6.8, from the rest frame, momentum $\mathbf{p} = 0$, to the frame in which the particle has energy E and momentum \mathbf{p} . The relevant spinors transform as²³

$$\cosh(\rho/2) = \frac{E+m}{\sqrt{2m(E+m)}} \quad \xi(\mathbf{p}) = [\cosh(\rho/2) - \mathbf{n}_\rho \cdot \boldsymbol{\sigma} \sinh(\rho/2)]\xi(0) \quad (6.13)$$

$$\sinh(\rho/2) = \frac{|\mathbf{p}|}{\sqrt{2m(E+m)}} \quad \eta(\mathbf{p}) = [\cosh(\rho/2) + \mathbf{n}_\rho \cdot \boldsymbol{\sigma} \sinh(\rho/2)]\eta(0) \quad (6.14)$$

which can be written as

$$\xi(\mathbf{p}) = \frac{E+m + \mathbf{p} \cdot \boldsymbol{\sigma}}{\sqrt{2m(E+m)}}\xi(0) \quad (6.15)$$

$$\eta(\mathbf{p}) = \frac{E+m - \mathbf{p} \cdot \boldsymbol{\sigma}}{\sqrt{2m(E+m)}}\eta(0) \quad (6.16)$$

²²P. A. M. Dirac, 1902–1984.

²³We have

In the particle's rest frame and in all frames moving with respect to it slowly enough that the Lorentz boost can be approximated by a Galilean transformation (not affecting time) when transforming between those frames, any differences in how spinors with dotted or undotted indexes transform disappear. In that case, spinors effectively live in 3 real dimensions. Inspecting eqn 6.5, we can see that in the limit $\beta \rightarrow 0$, L tends to the unit matrix and therefore, under the Galilean transformation, spinors do not change. Thus, at rest, both Weyl spinors, ξ^α and $\eta_{\dot{\beta}}$, become effectively identical to the same Pauli spinor and we can write $\xi^\alpha(0) = \eta_{\dot{\alpha}}(0)$. This allows us, after some algebra, to remove $\mathbf{p} = 0$ spinors from eqns 6.15 and 6.16 and to obtain the Dirac equation, eqn 6.12.

Thus the Dirac equation is equivalent to the Lorentz boost. This should be expected—once an object, like a bispinor, is found to represent a particle in its rest frame, the only thing left to do is to boost it to another frame as needed.

6.2 One-particle states

The fact that quantum states of free relativistic particles are fully defined by the Lorentz transformation supplemented by the space-time translation was discovered by Wigner.²⁴ Here we will follow his idea in a qualitative way just to get the main concept across.

First, we note that Lorentz transformations are not able to transform a given arbitrary 4-momentum p^μ into every possible p'^μ . Instead, the vector space of 4-momenta is divided into subspaces of 4-momenta that can be Lorentz-transformed into each other. Three of those subspaces represent experimentally known states. The simplest, at this stage, is the vacuum state given by the conditions $p^\mu = 0$ and $p^\mu p_\mu = 0$. There is no Lorentz transformation that would transform a 4-momentum not satisfying these conditions into one that does, and vice versa. We will not study vacuum states in this book, and therefore we move directly to consider two other possibilities.

A 4-momentum subspace related to massive particles, like the electron, is given by the condition $p^\mu p_\mu > 0$. In addition to a 4-momentum p^μ , what other degrees of freedom are present and which geometrical object represent them? To answer this question, we can consider Lorentz transformations that leave p^μ invariant.²⁵ To see what these are, we can transform p^μ to the particle rest frame, where $p'^\mu = (\text{mass}, 0, 0, 0)$, find the largest subset of Lorentz transformations leaving p'^μ invariant, and then transform back to the same p^μ . It turns out, as intuitively expected, that the desired transformations are space rotations acting on $2s + 1$ spinors representing $2s + 1$ spin projections of a spin- s particle. Thus, the electron, $s = \frac{1}{2}$, is represented by two Dirac spinors—in fact, by two Dirac spinors multiplied by a dimensionless scalar. To get the scalar, we add space–time translations. Looking for a theory that is space–time translation-invariant,²⁶ we are looking for the free-particle energy and momentum eigenstates that, in the position representation, lead to the scalar $\exp(-ip^\mu x_\mu)$.

²⁴Eugene Wigner, 1902–1995.

²⁵The group of such transformations is known as the little group.

²⁶Implying energy and momentum conservation.

The third 4-momentum subspace is defined by the conditions $p^\mu \neq 0$ and $p^\mu p_\mu = 0$. Photons belong to this class. The question is again to find the largest subset of Lorentz transformations leaving p^μ invariant. There is no rest frame in this case, and therefore, instead, we transform an arbitrary p^μ to the frame where $p'^\mu = (\omega, 0, 0, \omega)$. We can see that the largest²⁷ subset of the Lorentz transformations leaving p'^μ invariant are the rotations in the (x^1, x^2) plane. As a result, a spin- s massless particle is represented by only one state, a helicity eigenstate, and not by $2s + 1$ states as in the massive case. This is an important difference. In order to get parity-conserving electromagnetism with photons having either helicity $+$ or helicity $-$ states, we put those two, in principle different, helicity states into one theory.

²⁷For a discussion of some subtle issues, see [107].

6.2.1 Fields and probability amplitudes

We have now everything needed to develop RQM and to describe fundamental particles and their interactions. But before we move on, we pause to look at a larger picture of which RQM is only a part. The Dirac equation, for example, can be studied in the context of classical field theory or RQF or RQM. The algebra will often be identical, but the basic objects and the interpretation are different.

The most natural way to proceed from here would be to study a classical field theory. The paradigm for this is classical electromagnetism described in terms of the tensor field $F^{\mu\nu}$ or the 4-vector potential field A^μ :

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$$

Then such a field, for example a classical electron field, i.e. a classical Dirac spinor field Ψ , would be quantized, promoting the Ψ of classical field theory to an operator Ψ of RQF. Probability amplitudes would be obtained by taking matrix elements of Ψ of RQF sandwiched between particle states living in a suitably constructed space. In RQM, Ψ represents a particle state and is not an operator. Taking the vacuum to one-particle matrix element of the field operator Ψ of RQF, we get Ψ of RQM, which in the position representation is called the wavefunction. In RQM, one can also have states describing many particles—but only a fixed number of them. At high energies, much larger than the masses of the particles involved, particles can be created and the number of particles cannot be fixed; RQM is not adequate for this and RQF has to be used instead.

It is not appropriate in this text to go into sufficient detail to enable a proper understanding of classical field theory and RQF. Fortunately, considering the most important aspects of physics that we require, RQF gives the same results as those we will obtain in RQM. Differences will be in details beyond leading effects. The one important exception is that we will be missing the idea of a vacuum state. In RQF, a vacuum is not a ‘nothingness’, although particles are absent. For example, the

QCD vacuum is a very complicated state. We will, however, consider the nature of the electroweak vacuum in Chapter 12, since it is fundamental to the origin of mass in the Standard Model.

6.3 The Klein–Gordon equation

RQM of spin-0 particles was considered by Schrödinger first, before he published his famous equation for the non-relativistic case. He abandoned RQM because of formal difficulties that were only understood many years later. Here, we will see what they are and then define an area of applicability of RQM.

As argued earlier, a spin-0 particle with

$$p^\mu p_\mu = m^2 > 0 \quad (6.17)$$

in the position representation is expected to be described by a scalar wavefunction $\sim \exp(-ip^\mu x_\mu)$. Replacing the energy by $i\partial/\partial t$ and the momentum by $-i\nabla$ in eqn 6.17,²⁸ we get the Klein–Gordon (KG) equation²⁹ of RQM in the position representation:

$$(\square + m^2)\Psi(t, \mathbf{x}) = 0 \quad (6.18)$$

where

$$\square = \partial^\mu \partial_\mu = \frac{\partial^2}{\partial t^2} - \nabla^2$$

For a particle at rest, $-i\nabla\Psi(t, \mathbf{x}) = 0$, only the time (proper time τ) derivative would be present in eqn 6.18 and there would be two independent solutions: $\Psi^\pm(\tau, \mathbf{x}) = \exp(\mp im\tau)\Psi^\pm(0, 0)$. Therefore, in a frame in which the particle has momentum p and energy $E_p = +\sqrt{p^2 + m^2} > 0$ (the subscript ‘p’ is for the plus sign, indicating that E_p is positive), we get, as expected,³⁰

$$\Psi^+(t, \mathbf{x}) = N \exp(-ip \cdot x) = N \exp(-iE_p t + i\mathbf{p} \cdot \mathbf{x}) \quad (6.19)$$

where N is a normalization constant that will be defined shortly.

Instead of boosting the other solution, i.e. taking the $(-E_p, \mathbf{p})$ eigenstate, we take the complex conjugate of $\Psi^+(t, \mathbf{x})$, corresponding to the $(-E_p, -\mathbf{p})$ eigenstate, to get³¹

$$\Psi^-(t, \mathbf{x}) = N \exp(+ip \cdot x) = N \exp(+iE_p t - i\mathbf{p} \cdot \mathbf{x}) \quad (6.20)$$

By a direct substitution, one can check that a general solution of eqn 6.18 is indeed a linear combination of $\Psi^+(t, \mathbf{x})$ and $\Psi^-(t, \mathbf{x})$.

We have obtained, as expected, $\Psi^+(t, \mathbf{x})$, but, in addition, we also have $\Psi^-(t, \mathbf{x})$. This is the first puzzle of RQM, the nature of which will become clearer when we progress a little further. Both solutions of the KG equation are eigenfunctions of the energy operator $i\partial/\partial t$: $\Psi^+(t, \mathbf{x})$

²⁸We have

$$\begin{aligned} p^\mu &= i\partial^\mu \\ p^0 &= i\frac{\partial}{\partial x^0} = i\frac{\partial}{\partial t} \\ p^i &= i\partial^i = -i\partial_i = -i\frac{\partial}{\partial x^i} \end{aligned}$$

²⁹Sometimes known as the Klein–Gordon–Fock equation.

³⁰We have $m\tau = p^\mu x_\mu$.

³¹Why we are doing this should become clear after reading Section 6.3.1.

with an eigenvalue E_p and $\Psi^-(t, \mathbf{x})$ with $-E_p$, a negative energy for a free particle!

In exactly the same way as for the non-relativistic Schrödinger equation, we can derive the continuity equation for a probability density ρ and a probability current \mathbf{j} :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 \quad (6.21)$$

where

$$\rho = i \left(\Psi^* \frac{\partial \Psi}{\partial t} - \Psi \frac{\partial \Psi^*}{\partial t} \right) \quad (6.22)$$

$$\mathbf{j} = -i(\Psi^* \nabla \Psi - \Psi \nabla \Psi^*) \quad (6.23)$$

The probability current turns out to be given by the same expression as in the non-relativistic case, but the probability density is different, although it shows a nice symmetry with the current, and we can define a 4-vector current

$$j^\mu \equiv (\rho, \mathbf{j}) = i(\Psi^* \partial^\mu \Psi - \Psi \partial^\mu \Psi^*)$$

The continuity equation, eqn 6.21, can be then written as $\partial_\mu j^\mu = 0$. The corresponding conserved quantity is the total probability, which we obtain by integrating $j^0 = \rho$ over the 3-dimensional space. The underlying symmetry is invariance with respect to multiplication by a global phase factor: physics described by Ψ is identical to physics described by $e^{i\theta} \Psi$ for any fixed real parameter θ .³²

Substituting $\Psi^+(t, \mathbf{x})$ from eqn 6.19 into eqn 6.23, we obtain

$$\rho^+ = 2|N|^2 E_p, \quad \mathbf{j}^+ = 2|N|^2 \mathbf{p} \quad (6.24)$$

Now we can fix the normalization N . In NRQM, the volume integral of the probability density is a constant with value 1 for one particle in the whole space. This does not work in RQM, because of the Lorentz contraction, which modifies the volume, contracting one side of a cube, parallel to the Lorentz boost, by the Lorentz factor γ . To keep the integral independent of the Lorentz transformation, the probability density should grow by the same factor³³ γ . So putting $N = 1$ would do the job, as would any other constant. The choice of $N = 1$ is called the covariant normalization and corresponds to $2E_p$ particles in a unit volume. Another popular choice is $N = 1/\sqrt{2m}$, which in the non-relativistic limit $E_p \rightarrow m$ makes $\rho \rightarrow \Psi^* \Psi$ and $\mathbf{j} \rightarrow \mathbf{velocity}$ approach the expressions from NRQM.

For $\Psi^-(t, \mathbf{x})$, eqn 6.24 becomes

$$\rho^- = -2|N|^2 E_p, \quad \mathbf{j}^- = -2|N|^2 \mathbf{p} \quad (6.25)$$

In summary, $\Psi^+(t, \mathbf{x})$ and related observables, the energy, the probability density, and the probability current come out as expected and

³²This can be proved from the Klein-Gordon Lagrangian and Noether's theorem.

³³There is no problem here, since ρ is the time-like component of a 4-vector.

behave nicely in the non-relativistic limit. In contrast to $\Psi^+(t, \mathbf{x})$, an unexpected additional wavefunction $\Psi^-(t, \mathbf{x})$ describes a free particle with negative energy and negative probability density and with the probability current flowing in the opposite direction to the particle's momentum—all properties that are unexpected and difficult to accept.

6.3.1 The Feynman–Stueckelberg interpretation of negative-energy states

In this section, we outline the Feynman–Stueckelberg interpretation of negative-energy states following the approach described by Feynman in his Dirac Lecture [74], which is recommended as further reading.

Suppose there is a particle in a state ϕ_0 as indicated in Fig. 6.2(a). At time t_1 , a potential U_1 is turned on for a moment, acting on the particle and changing its state to an intermediate state. At time t_2 , a second perturbation U_2 changes that intermediate state to the final one, which could be the same as the original state ϕ_0 . The amplitude for the particle to go from the initial state ϕ_0 to the same state ϕ_0 after time t_2 has a contribution from the amplitude with an intermediate state, existing for the period of time from t_1 to t_2 , of energy $E_p > 0$. All possible intermediate states of different energies $E_p > 0$ contribute. Among them, there are amplitudes for particles travelling faster than the speed of light. This is the result of insisting that all energies E_p be positive. If one starts a series of waves from a point, keeping all energies positive, these waves cannot be confined to the inside of the light cone. The sketch in Fig. 6.2(a) corresponds to an amplitude where the particle in the intermediate state travels faster than the speed of light. An observer in the reference frame (a) with coordinates (t, x) observes one particle in quantum state ϕ_0 that moves from x_1 to x_2 , ending in the same quantum state ϕ_0 .

As indicated in Fig. 6.2(b), there is another reference frame (b) with coordinates (t', x') in which the sequence of events is different; t'_2 happens first, before t'_1 . An observer in this reference frame has a different story to tell. A particle at x'_1 is in a quantum state ϕ_0 . Nothing happens until time t'_2 , when suddenly two particles emerge from the point x'_2 . One of these particles travels to x'_1 and at time t'_1 collides with the original particle. The particles annihilate with each other, disappearing

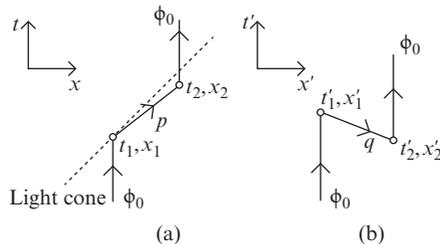


Fig. 6.2 A contribution to the transition amplitude viewed in two different reference frames. Adapted from [74].

from the scene, leaving the third particle at x'_2 in state ϕ_0 . In frame (b), three particles were present between t'_2 and t'_1 . The second observer can argue that the particle that travelled from x'_2 to x'_1 is the antiparticle of the original particle and therefore they were able to annihilate with each other. So antiparticles must exist and their properties are defined from particles such that the annihilation works. The first observer might argue that the antiparticle of the second observer is her/his particle travelling backwards in time and that is the interpretation of the negative-energy states. The negative-energy states correspond to particles travelling backwards in time and therefore the phase of the wavefunction in eqn 6.20 has a $-iE_p(-t) \equiv +iE_p t$ contribution instead of $-iE_p t$ as in eqn 6.19. To make the picture complete, we must also take into account that a particle travelling backwards in time has its momentum reversed. Mathematically, all this is equivalent to taking the complex conjugate of the positive-energy solution Ψ^+ (eqn 6.19) to obtain the negative-energy solution Ψ^- (eqn 6.20).

In summary, negative-energy solutions of the KG equation represent antiparticles. The probability density represents the charge density and can be either negative or positive. The same applies for the probability current, representing the charged current, the number of charges passing through the unit area per unit time.

Inclusion of interactions via a potential

We introduce interactions using a potential and see what happens. Following the way in which a potential V is introduced into the non-relativistic Schrödinger eqn, we modify the energy operator³⁴

$$i\frac{\partial}{\partial t} \rightarrow i\frac{\partial}{\partial t} - V$$

transforming eqn 6.18 into

$$\left(i\frac{\partial}{\partial t} - V\right)^2 \Psi = (-\nabla^2 + m^2)\Psi$$

which, for a time-independent, time-like potential V in one dimension and for energy eigenstates with energy E_p , becomes³⁵

$$\begin{aligned} i\frac{\partial \Psi}{\partial t} &= E_p \Psi \\ \Psi(t, s) &= \psi(s) \exp(-iE_p t) \end{aligned} \quad [E_p - V(s)]^2 \psi(s) = \left(-\frac{\partial^2}{\partial s^2} + m^2\right) \psi(s) \quad (6.26)$$

Consider a time-like potential barrier of fixed height $V > 0$ for $s \geq 0$, as shown in Fig. 6.3. The wavefunction $\psi(s)$ consists of incident, Ie^{ips} , reflected, Re^{-ips} , and transmitted, Te^{iks} , waves:

$$\begin{aligned} \psi_L(s) &= Ie^{ips} + Re^{-ips} \\ \psi_R(s) &= Te^{iks} \end{aligned}$$

³⁴A proper discussion of interactions will be given in Section 6.5.

³⁵We have

where

$$\psi(s) = \begin{cases} \psi_L(s) & \text{for } s < 0 \\ \psi_R(s) & \text{for } s \geq 0 \end{cases}$$

On substituting ψ_L and ψ_R into eqn 6.26, we obtain $E^2 = p^2 + m^2$ and $(E - V)^2 = k^2 + m^2$, leading to

$$p = \pm\sqrt{E^2 - m^2}, \quad k = \pm\sqrt{(E - V)^2 - m^2}$$

In both cases, we choose a + sign in front of the square root to match the expected propagation directions as in Fig. 6.3.

From the continuity condition at $s = 0$ for $\psi(s)$ and $d\psi(s)/ds$, we get

$$I + R = T, \quad pI - pR = kT$$

and so

$$T = \frac{2p}{p+k}I, \quad R = \frac{p-k}{p+k}I \quad (6.27)$$

The probability currents along s for $s < 0$ and $s \geq 0$ are

$$j_L = \frac{p}{m}(|I|^2 - |R|^2), \quad j_R = \frac{k}{m}|T|^2 \quad (6.28)$$

Keeping the energy E fixed, we consider three different cases of potential strength.

The first case is that of a weak potential, $E > V + m$, where k is real and $k < p$. The probability densities in the two regions are

$$\rho_L = \frac{E}{m}|\psi_L|^2 > 0, \quad \rho_R = \frac{E - V}{m}|\psi_R|^2 > 0 \quad (6.29)$$

This case looks like the non-relativistic one, with nothing special happening: a small fraction of the incoming wave is reflected and the rest is transmitted.

In the second case, the potential is of moderate strength, $V - m < E < V + m$, and $k = i\sqrt{m^2 - (E - V)^2} = i\kappa$ is purely imaginary (with κ real) and

$$R = \frac{p - i\kappa}{p + i\kappa}I$$

so

$$|R| = |I|, \quad j_L = 0$$

The incoming wave is totally reflected and the probability density in the barrier shows the expected exponential decay

$$\rho_R = \frac{E - V}{m}|\psi_R|^2 = \frac{E - V}{m}e^{-2\kappa s}$$

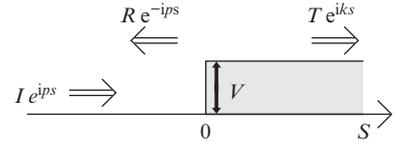


Fig. 6.3 A time-like potential barrier of height V . Incoming, reflected, and transmitted waves are also indicated.

as in the non-relativistic case. However, the situation is not identical, because the increasing potential V changes the sign of the probability density from positive ($\rho_R > 0$ in the first case of the weak potential) to negative:

$$E > V \quad \Rightarrow \quad \rho_R > 0$$

but

$$E < V \quad \Rightarrow \quad \rho_R < 0 \tag{6.30}$$

We will come back to this after discussing the case of the strong potential, $E < V - m$, when k becomes purely real and $k^2 > p^2$. The probability, which became negative when the potential became strong enough in the previous case, stays negative and the probability current becomes real inside the barrier:

$$\rho_R = \frac{E - V}{m} |T|^2 < 0, \quad j_R = \frac{k}{m} |T|^2$$

Consider next the previously unphysical case of $k < 0$, because, in a counter-intuitive way, this case now corresponds to a particle moving to the right. To see that, we will calculate the group velocity

$$\mathcal{V}_g = \frac{\partial E}{\partial k} = \frac{k}{E - V} > 0$$

A consequence is that T and R given by eqn 6.27 can be arbitrarily large, possibly making the reflected current bigger than the incoming one.³⁶ The reason for this is that the very strong potential provides enough energy to produce particle–antiparticle pairs. The probability current and the probability density within the barrier are negative because created antiparticles (see eqn 6.25) are attracted to the barrier, moving to the right ($\mathcal{V}_g > 0$). The created particles are repelled by the barrier, moving to the left, increasing the reflected probability current. Such a situation can be created by focusing light from a high-power laser, making a very strong electric field, which in turn produces electron–positron pairs from the vacuum. Another example is the Hawking radiation in the neighbourhood of a black hole. The fact that the probability density already became negative in the case of a moderate-strength potential corresponds to vacuum polarization by the creation of virtual particle–antiparticle pairs. They do not affect the probability currents, because there is not enough energy in the system to promote them to become real particles. An analogy is the Lamb shift in atomic physics, where the vacuum polarization affects energy levels. The problem of RQM is now clear: the formalism describes one particle (or a fixed number of particles), but physics needs many particles, the number of which cannot be fixed—particles can be created and particles can be annihilated. One needs RQF to describe such physics. RQM can only be used as long as the number of particles is fixed. Using the uncertainty relation

³⁶This is known as the Klein paradox, another puzzle of RQM, after O. Klein, 1894–1977.

$\Delta p \Delta s \sim \hbar$, we see that pair creation that starts at $\Delta p \sim mc$ sets the limit on $\Delta s \sim \hbar/mc$. So, as long as we are studying physics at a scale bigger than \hbar/mc , known as the Compton wavelength, RQM can be applied. Atomic physics is an example where this condition is usually fulfilled: the Compton wavelength of the electron is about 3.86×10^{-13} m. But RQM can also be applied to many processes in high-energy particle physics. A representative example is electron–positron annihilation producing hadrons, many of which are pions. The fundamental process in this case is $e^+ + e^- \rightarrow q + \bar{q}$. The number of particles is 2 and is fixed, and the change from 2 leptons to 2 quarks can be handled by RQM. Fragmentation of quarks to hadrons takes place on a different, much slower, time scale and therefore can be separated from the fundamental process of $e^+ + e^-$ annihilation.

Is there any limit on Δp ? How well can one measure momentum? In NRQM, momentum can be measured with any precision, but, because of the upper limit on speed, $< c$, we must have³⁷

$$\Delta p \Delta t \sim \frac{\hbar}{c}$$

and consequently infinite precision $\Delta p \rightarrow 0$ requires infinite measurement time $t \rightarrow \infty$.

³⁷See the Introduction in [50].

6.4 The Dirac equation

This is the most important section of this chapter. The Dirac equation provides a relativistically consistent equation describing a massive point-like spin- $\frac{1}{2}$ particle such as the electron and it led to the prediction of the positron—the first antiparticle.

First, we consider different representations of the Dirac equation, the probability current and bilinear covariants. Then we find the free-particle states, examine their properties, and introduce chirality and helicity operators. The formalism is then applied to describe simple Standard Model processes such as $e^+ + e^- \rightarrow \mu^+ + \mu^-$ at energies well above the muon rest mass but well below the Z^0 mass.³⁸

The properties of Dirac particles under the discrete symmetries P , C , and T are then discussed. Electromagnetic interactions are introduced via so-called minimal coupling and the non-relativistic limit is obtained, leading to the prediction of $g = 2$ for the magnetic dipole moment of the electron. Finally, there is a brief discussion of the Aharonov–Bohm effect and the pre-eminence of the electromagnetic 4-vector potential in RQM.

Before we move on, we shall give a few words of introduction for readers who skipped Section 6.1. Unlike the non-relativistic case, where electron spin is described by a column of two complex numbers, called the Pauli spinor,³⁹ in RQM one needs a pair of two-component spinors ξ^α and $\eta_{\dot{\beta}}$ (with $\alpha = 1, 2$ and $\dot{\beta} = 1, 2$) called the undotted and dotted spinors (the latter being distinguished by a dot above the index)—these

³⁸The domain of the e^+e^- colliders PETRA (DESY) and PEP (SLAC).

³⁹Following experimental observations suggesting that the electron has a property called spin, Pauli extended the Schrödinger equation describing the electron's interaction with an electromagnetic field by inserting into it a two-component spinor and corresponding magnetic dipole moment.

are just names, and they could be called ‘apples’ and ‘pears’ since they are as different as apples and pears. The reason for the different names is that the corresponding spinors transform differently under Lorentz transformations (see eqns 6.7 and 6.8). Combining ξ^α and $\eta_{\dot{\beta}}$ into one

4-component column gives the Dirac spinor Ψ : $\Psi = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$. Now this Dirac spinor can be transformed using a unitary operator (changing what is called a representation), mixing ξ and η . And we must follow these different ξ and η spinors all the way through to know how to Lorentz-transform the resulting Dirac spinor; having just the four complex numbers constituting the Dirac spinor is not enough to know how to apply the Lorentz transformation to them.

The Dirac equation, as derived in Section 6.1 as eqn 6.12 or, for readers who skipped that section, as derived in Exercises 6.2 and 6.3 following the historical path, can be written as

$$\begin{pmatrix} 0 & p_0 + \mathbf{p} \cdot \boldsymbol{\sigma} \\ p_0 - \mathbf{p} \cdot \boldsymbol{\sigma} & 0 \end{pmatrix} \Psi = m\Psi \quad (6.31)$$

Instead of Ψ , we could use $\Psi' = U\Psi$, where U is a unitary operator. In the new basis, the new representation, eqn 6.31, would look different. In general, the Dirac equation can be written as

$$(\gamma p - m)\Psi = 0 \quad (6.32)$$

where

$$\gamma p \equiv \gamma^\mu p_\mu = p_0 \gamma^0 - \mathbf{p} \cdot \boldsymbol{\gamma} = i\gamma^0 \frac{\partial}{\partial t} + i\boldsymbol{\gamma} \cdot \nabla \quad (6.33)$$

⁴⁰In some books, ξ^α and $\eta_{\dot{\beta}}$ swap places, leading to different space-like $\boldsymbol{\gamma}$ matrices, multiplied by -1 .

Comparing equations 6.32 and 6.31, we can see that in the representation that was used to get eqn 6.31,

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 0 & -\boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \quad (6.34)$$

⁴¹We have

$$\begin{aligned} (\gamma^0)^2 &= 1 \\ (\gamma^1)^2 &= (\gamma^2)^2 = (\gamma^3)^2 = -1 \\ \gamma^i &= U\gamma U^\dagger = U\gamma U^{-1} \end{aligned}$$

This is known as the Weyl or symmetric or chiral representation.⁴⁰ Multiplying eqn 6.32 by γp from the left, we get representation-independent constraints on the γ matrices:

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu} \quad (6.35)$$

⁴²Equation 6.32 is

$$i\gamma^0 \frac{\partial \Psi}{\partial t} + i\boldsymbol{\gamma}^k \frac{\partial \Psi}{\partial x^k} - m\Psi = 0$$

The matrix γ^0 is Hermitian and the matrices γ^i are anti-Hermitian (in any representation):⁴¹

Applying \dagger , we get

$$i\frac{\partial \Psi^\dagger}{\partial t} \gamma^0 + i\frac{\partial \Psi^\dagger}{\partial x^k} (-\gamma^k) + m\Psi^\dagger = 0$$

$$\gamma^{0\dagger} = \gamma^0, \quad (\gamma^i)^\dagger = -\gamma^i \quad (6.36)$$

and multiplying by γ^0 from the right and using eqn 6.35 gives

Applying Hermitian conjugation to eqn 6.32, using properties of the γ matrices given by eqn 6.36 and after some algebra,⁴² we get the adjoint Dirac equation

$$i\frac{\partial \bar{\Psi}}{\partial t} \gamma^0 + i\frac{\partial \bar{\Psi}}{\partial x^k} \gamma^k + m\bar{\Psi} = 0$$

$$\bar{\Psi}(\gamma p + m) = 0 \quad (6.37)$$

where the adjoint spinor is

$$\bar{\Psi} \equiv \Psi^\dagger \gamma^0 \quad (6.38)$$

and p acts on the left.

Multiplying eqn 6.32 by $\bar{\Psi}$ from the left and eqn 6.37 by Ψ from the right and adding the resulting equations, we get

$$\bar{\Psi} \gamma^\mu \partial_\mu \Psi + (\partial_\mu \bar{\Psi}) \gamma^\mu \Psi = \partial_\mu (\bar{\Psi} \gamma^\mu \Psi) = 0$$

which is the continuity equation, $\partial_\mu j^\mu = 0$, for the probability current 4-vector

$$j^\mu = \bar{\Psi} \gamma^\mu \Psi \quad (6.39)$$

The probability density

$$\rho \equiv j^0 = \bar{\Psi} \gamma^0 \Psi = \sum_{i=1}^4 |\Psi_i|^2 \quad (6.40)$$

is the time-like component of the probability current; it is positive-definite and has a similar form to the non-relativistic expression. Ψ and $\bar{\Psi}$ may be used to form quantities with well-defined space-time transformation properties—known as bilinear covariants. The simplest is the Lorentz scalar $\bar{\Psi} \Psi$.⁴³ The next simplest is $\bar{\Psi} \gamma^\mu \Psi$, which transforms as a Lorentz 4-vector.

The Hamiltonian H of the Dirac equation is obtained by multiplying eqn 6.32 by γ^0 from the left and separating the time derivative:

$$H\Psi = i \frac{\partial \Psi}{\partial t} \quad (6.41)$$

where

$$H = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m \quad (6.42)$$

and

$$\boldsymbol{\alpha} = \gamma^0 \boldsymbol{\gamma}, \quad \beta = \gamma^0 \quad (6.43)$$

The matrices $\boldsymbol{\alpha}$ and β are Hermitian and in the Weyl representation are given by⁴⁴

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & -\boldsymbol{\sigma} \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (6.44)$$

The Weyl representation is very well suited to the ultrarelativistic limit, where the mass can be neglected, because the Dirac bispinor is then effectively reduced to a single Weyl spinor. In the non-relativistic limit, however, both Weyl spinor components of the Dirac bispinor contribute

⁴³The matrix γ^0 that is inside $\bar{\Psi}$ swaps spinors in the bispinor such that the dotted index meets the dotted one ($\xi^* \eta$) and the undotted index meets the undotted one ($\eta^* \xi$). It should be noted that complex conjugation adds or removes a dot, so if ξ has an undotted index, ξ^* has a dotted index.

⁴⁴We have

$$\begin{aligned} \alpha_i \alpha_j + \alpha_j \alpha_i &= 2\delta_{ij} \\ \beta \boldsymbol{\alpha} + \boldsymbol{\alpha} \beta &= 0 \\ \beta^2 &= 1 \end{aligned}$$

equally, so another representation, called the standard or Dirac representation, is more suitable. This representation is often used in introductory textbooks. The transformation from the Weyl representation to the Dirac representation is effected by the unitary transformation

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

which gives

$$\Psi(\text{Dirac}) = \begin{pmatrix} \varphi \\ \chi \end{pmatrix} = U\Psi(\text{Weyl}) = U \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \xi + \eta \\ \xi - \eta \end{pmatrix} \quad (6.45)$$

The transformation of the γ matrices, $\gamma(\text{Dirac}) = U\gamma(\text{Weyl})U^{-1}$, gives

$$\gamma^0 = \beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}$$

The Dirac equation in the (standard) Dirac representation is then given by

$$\begin{aligned} E\varphi - \mathbf{p} \cdot \boldsymbol{\sigma}\chi &= m\varphi \\ -E\chi + \mathbf{p} \cdot \boldsymbol{\sigma}\varphi &= m\chi \end{aligned} \quad (6.46)$$

In the non-relativistic limit, $\chi \rightarrow 0$ and the Dirac spinor becomes effectively a two-component Pauli spinor.

The fact that in the relativistic theory one needs two Weyl spinors to describe the electron while in the non-relativistic world one Pauli spinor is enough is always difficult to accept. To give some insight into why this is the case, consider yet another representation, that of Foldy and Wouthuysen (FW). We start with the Dirac representation and apply a momentum-dependent unitary transformation U_{FW} given by

$$U_{\text{FW}} = \exp\left(\frac{1}{2} \frac{\beta \boldsymbol{\alpha} \cdot \mathbf{p}}{|\mathbf{p}|} \arctan \frac{|\mathbf{p}|}{m}\right)$$

The wavefunction in the FW representation is then

$$\Psi(\text{FW}) = \begin{pmatrix} u \\ w \end{pmatrix} = U_{\text{FW}}\Psi(\text{Dirac}) = U_{\text{FW}} \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$$

and after the transformation of the Hamiltonian, eqn 6.42 splits into components, becoming

$$\sqrt{\mathbf{p}^2 + m^2} u = i \frac{\partial u}{\partial t} \quad (6.47)$$

$$-\sqrt{\mathbf{p}^2 + m^2} w = i \frac{\partial w}{\partial t} \quad (6.48)$$

We now have two decoupled equations for positive- and negative-energy solutions, respectively. But if we want to drop one of them and consider, say, only the equation for the positive energies, then there is a problem, because we would not know how to transform the positive-energy spinor without any knowledge of the other one.⁴⁵ In the non-relativistic limit, however, $\sqrt{\mathbf{p}^2 + m^2} \simeq m + |\mathbf{p}|^2/2m$, which leads to the Schrödinger Hamiltonian, and the Lorentz transformation becomes a Galilean one, which does not affect spin. So, in the non-relativistic limit, we can just take one equation, for example the one for the positive energy, and use it to describe a non-relativistic electron, with its spinor u being effectively the Pauli spinor. We can ‘forget’ about the negative-energy solution.

⁴⁵We only know how to transform ξ^α and $\eta_{\dot{\beta}}$ and they are buried inside the u and w spinors. We would need to get ξ^α and $\eta_{\dot{\beta}}$ out of u and w , transform u and w back to ξ^α and $\eta_{\dot{\beta}}$, do the Lorentz transformation, and transform back to get the Lorentz-transformed u and w .

Majorana particles

This is a short, rather technical, detour from the main track to introduce the concept of the Majorana particle.⁴⁶ It is not essential for what follows. No fundamental Majorana particle has yet been discovered, although it could be that neutrinos are Majorana particles, and a composite Majorana particle has been discovered in condensed matter. From the discussion of spinors in Section 6.1.1, one might get the impression that a massless spin- $\frac{1}{2}$ particle is described by the Weyl spinor and a massive one by the Dirac spinor, which has two independent Weyl spinors as components. Although all the particles we know either fit, or could fit, this scenario, this is not the only scenario. We can start with one Weyl spinor, say, an undotted spinor ξ^α , get a dotted one by complex conjugation, and then lower the index with the metric tensor ϵ to give a spinor that transforms like $\eta_{\dot{\beta}}$. We can then use these ξ - and η -like objects to construct a Dirac spinor that satisfies the Dirac equation in the Weyl representation for a massive particle, the Majorana particle, effectively defined by one Weyl spinor rather than by two independent Weyl spinors as for the electron.

⁴⁶Named after Ettore Majorana, 1906–1938.

6.4.1 Free-particle solutions

The Dirac equation, eqn 6.46, in the Dirac (standard) representation can be written as

$$\begin{pmatrix} m & \mathbf{p} \cdot \boldsymbol{\sigma} \\ \mathbf{p} \cdot \boldsymbol{\sigma} & -m \end{pmatrix} \begin{pmatrix} \varphi \\ \chi \end{pmatrix} = E \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$$

which for an electron at rest simplifies to

$$\begin{pmatrix} m & 0 \\ 0 & -m \end{pmatrix} \begin{pmatrix} \varphi \\ \chi \end{pmatrix} = i \frac{\partial}{\partial t} \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$$

There are four independent, un-normalized, solutions for the energy eigenstates:

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-im\tau}, & \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} e^{-im\tau} & \quad \text{for } E = m \\ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{+im\tau}, & \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} e^{+im\tau} & \quad \text{for } E = -m \end{aligned}$$

⁴⁷We can go back to the Weyl representation, boost ξ (eqn 6.15) and η (eqn 6.16), and return to the Dirac representation.

By boosting these solutions to the frame where $\mathbf{p} \neq 0$, we get ($s = 1, 2$)⁴⁷

$$u^{(s)} \exp[-i(E_p t - \mathbf{p} \cdot \mathbf{x})] = u^{(s)} \exp(-ip \cdot x) \quad \text{for } E > 0$$

and, for the $-E_p$ energy eigenstate and \mathbf{p} momentum eigenstate,

$$u^{(s+2)} \exp(+iE_p t) \exp(i\mathbf{p} \cdot \mathbf{x}) = u^{(s+2)} \exp[+i(+E_p t + \mathbf{p} \cdot \mathbf{x})] \quad \text{for } E < 0$$

where

$$u^{(s)} = N \begin{pmatrix} \vartheta^{(s)} \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m} \vartheta^{(s)} \end{pmatrix}, \quad u^{(s+2)} = N \begin{pmatrix} \frac{-\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m} \vartheta^{(s)} \\ \vartheta^{(s)} \end{pmatrix}$$

N is a normalization constant, $E_p = +\sqrt{\mathbf{p}^2 + m^2}$, and

$$\vartheta^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vartheta^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

⁴⁸In principle, we could carry on with them, as is done in a number of textbooks.

As in the case of the KG equation, instead of the above two solutions for $E < 0$ energy eigenstates,⁴⁸ we will follow the Feynman–Stueckelberg interpretation of negative-energy states as antiparticles that are equivalent to particles travelling backwards in time. This requires us to replace the momentum \mathbf{p} by $-\mathbf{p}$ and to decide which spin state to choose in going from a particle to an antiparticle.

Taken together, these steps give the four independent solutions of the Dirac equation:

$$\Psi^+(x) = u^{(s)} e^{-ip \cdot x}, \quad \Psi^-(x) = v^{(s)} e^{+ip \cdot x} \quad (6.49)$$

where $u^{(s)}$ is as above and

$$v^{(1)}(\mathbf{p}) = u^{(4)}(-\mathbf{p}), \quad v^{(2)}(\mathbf{p}) = u^{(3)}(-\mathbf{p}) \quad (6.50)$$

Both Ψ^+ , describing a free electron, and Ψ^- , describing a free positron, are twofold-degenerate. We need a quantum number (label) to

distinguish the pairs of states with the same energy. A suitable operator, commuting with the free-particle Hamiltonian, is the helicity operator

$$h(\mathbf{p}) = \begin{pmatrix} \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} & 0 \\ 0 & \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \end{pmatrix} \quad (6.51)$$

In a convenient reference frame where $\mathbf{p} = (0, 0, p)$,

$$h(\mathbf{p}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and, for example,

$$u^{(1)} = N \begin{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \frac{p}{E_p + m} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix}$$

Thus $u^{(s)}$ (as well as $v^{(s)}$) are helicity eigenstates:

$$h(\mathbf{p})u^{(1)} = +u^{(1)}, \quad h(\mathbf{p})u^{(2)} = -u^{(2)}$$

Helicity eigenvalues correspond to the spin component along the direction of motion. Helicity + means that, in its rest frame, the electron has $+\frac{1}{2}$ spin projection on the axis parallel to \mathbf{p} (likewise for helicity – and the $-\frac{1}{2}$ spin projection); $\vartheta^{(1)}$ and $\vartheta^{(2)}$ are also defined with respect to this axis, which has to be the z axis, given the chosen representation of the Pauli matrices.

We can see now that changing \mathbf{p} to $-\mathbf{p}$ changes the direction of the quantization axis in the rest frame of the electron and the spin projection changes sign as in eqn 6.50 (see the s label).⁴⁹

⁴⁹See [13] for further reading.

As for the KG equation, we use covariant normalization, requiring that the integral of the probability density over the unit volume gives $2E_p$ particles. This gives the normalization constant as $N = \sqrt{E_p + m}$.

6.4.2 Chirality \neq helicity

Chirality, also called handedness, is defined by a pair of projection operators⁵⁰

$$P_L = \frac{1 - \gamma^5}{2}, \quad P_R = \frac{1 + \gamma^5}{2}, \quad \text{where } \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 \quad (6.52)$$

They divide the space of wavefunctions into right-handed and left-handed half-spaces. P_R projects a wavefunction onto the right-handed

⁵⁰These satisfy

$$\begin{aligned} P_L + P_R &= 1 \\ P_L P_R &= P_R P_L = 0 \\ P_L P_L &= P_L, \quad P_R P_R = P_R \end{aligned}$$

half-space, giving the right-handed component of the wavefunction. P_L does the same with respect to the left-handed half-space and the left-handed component of the wavefunction.⁵¹ Describing the Dirac spinor in the Dirac representation (ignoring the normalization) in terms of the Weyl spinors ξ and η of the Weyl representation (see eqn 6.45), we get

⁵¹Thus $\Psi = P_R \Psi + P_L \Psi = \Psi_R + \Psi_L$.

$$\frac{1 - \gamma^5}{2} \begin{pmatrix} \xi + \eta \\ \xi - \eta \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \xi + \eta \\ \xi - \eta \end{pmatrix} = \begin{pmatrix} \eta \\ -\eta \end{pmatrix}$$

$$\frac{1 + \gamma^5}{2} \begin{pmatrix} \xi + \eta \\ \xi - \eta \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \xi + \eta \\ \xi - \eta \end{pmatrix} = \begin{pmatrix} \xi \\ \xi \end{pmatrix}$$

We can thus define left-handed and right-handed spinors as

$$\eta_L \equiv \eta, \quad \xi_R \equiv \xi \quad (6.53)$$

⁵²The operators P_L and P_R respectively pull the dotted and undotted components out of the Dirac spinor.

The projection operators P_L and P_R are important in particle physics because weak interactions are sensitive to these components.⁵² The W boson couples only to the left-handed part of a particle wavefunction and the Z boson couples to both parts but with different strengths.

In order to gain better insight, we will consider chiral components of the Dirac spinor u :

$$\frac{1 - \gamma^5}{2} u = \frac{1 - \gamma^5}{2} \begin{pmatrix} \vartheta \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m} \vartheta \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m}\right) \vartheta \\ - \left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m}\right) \vartheta \end{pmatrix}$$

where $\vartheta = \vartheta^+ + \vartheta^-$ is a superposition of helicity $+$ (ϑ^+) and helicity $-$ (ϑ^-) eigenstates (not normalized). Because

$$\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \vartheta^+ = \vartheta^+, \quad \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \vartheta^- = -\vartheta^-$$

it follows that

$$\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta^+ = 0, \quad \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta^- = -2\vartheta^-$$

$$\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \left(1 + \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta^+ = 2\vartheta^+, \quad \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|} \left(1 + \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta^- = 0.$$

Finally,

$$\left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m}\right) \vartheta = \frac{a}{2} \left(1 + \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta + \frac{b}{2} \left(1 - \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|\mathbf{p}|}\right) \vartheta$$

where

$$a = 1 - \frac{|\mathbf{p}|}{E_p + m}, \quad b = 1 + \frac{|\mathbf{p}|}{E_p + m}$$

gives a decomposition of the final state into helicity + and helicity – components with weights $a/2$ and $b/2$, respectively. In summary,

$$\frac{1 - \gamma^5}{2} u = \frac{a}{2}(\text{helicity } +) + \frac{b}{2}(\text{helicity } -) = u_L \quad (\text{left chiral state})$$

$$\frac{1 + \gamma^5}{2} u = \frac{b}{2}(\text{helicity } +) + \frac{a}{2}(\text{helicity } -) = u_R \quad (\text{right chiral state})$$

It should be clear now that chirality and helicity are two different things. However, in the limit

$$\text{speed} \rightarrow c \quad \Rightarrow \quad a \rightarrow 0 \quad \text{and} \quad b \rightarrow 2$$

chirality and helicity become identical, which is the subject of the next section. Before that, we will explore consequences, relevant at energies where masses cannot be neglected, of chirality and helicity being different—for example affecting weak decay rates of particles. As an example, we will consider $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$ decay.

In the rest frame of the π^- , the momenta of μ^- and $\bar{\nu}_\mu$ are back to back and the helicities are as indicated in Fig. 6.4. The $\bar{\nu}_\mu$ is in the helicity + state,⁵³ with its spin along its momentum. The μ^- has to be in the helicity + state to conserve total angular momentum, since the pion has spin zero. But the W^- couples to the left-handed component of the μ^- wavefunction and therefore the μ^- needs, simultaneously, to have helicity + (to conserve angular momentum) and left-handed chirality (for the W^- to couple). The probability for this is

$$\frac{|a|^2}{|a|^2 + |b|^2} = \frac{1}{2} \left(1 - \frac{\text{speed}}{c} \right)$$

We can see that as the speed tends to c , the decay rate tends to 0, so the π^- cannot decay to a massless μ^- . This explains why, although favoured by the energy phase-space factor (not considered here), the decay rate for $\pi^- \rightarrow e^- + \bar{\nu}_e$ is much smaller than that for $\mu^- + \bar{\nu}_\mu$.⁵⁴

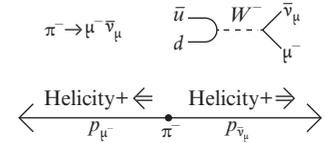


Fig. 6.4 $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$ decay.

⁵³We are not going to consider the helicity – state, since such a state has never been observed. A neutrino with non-zero mass is either described by a Dirac spinor with two Weyl spinors contributing but with one sterile helicity state or as a massive Majorana particle having only one helicity state.

⁵⁴See Exercise 6.4.

6.4.3 Helicity conservation and interactions via currents

It is worth repeating the conclusion of the previous section that chirality (handedness) is different from helicity and that this has consequences at low energies where masses cannot be neglected. At high energies, where masses can be neglected, helicity and chirality can be treated as identical and in many textbooks this is the working assumption right from the start.

It can be shown (see e.g. [84]) that in the probability current

$$j^\mu = \bar{u} \gamma^\mu u = (\bar{u}_L + \bar{u}_R) \gamma^\mu (u_L + u_R) = \bar{u}_L \gamma^\mu u_L + \bar{u}_R \gamma^\mu u_R$$

no cross terms like $\bar{u}_L \gamma^\mu u_R$ are present. In the high-energy limit where masses can be neglected,

$$\begin{aligned} \frac{1}{2}(1 - \gamma^5)u &= u_L \simeq u_L^- && \text{is the helicity } - \text{ eigenstate} \\ \frac{1}{2}(1 + \gamma^5)u &= u_R \simeq u_R^+ && \text{is the helicity } + \text{ eigenstate} \end{aligned}$$

and therefore the probability current does not contain any helicity-mixing terms like $\bar{u}_L^- \gamma^\mu u_R^+$. We will now discuss the significance of this.

In classical electromagnetism, two parallel wires carrying electric currents interact with each other with a force proportional to the product of the currents. We can try a similar idea to describe scattering of particles. Multiplying the probability current of a free electron by its electric charge will give us a 4-vector with time-like component representing the charge density and space-like component representing the number of electric charges crossing unit area per unit time, i.e. the electric current corresponding to moving electrons. If we now take that current and another one, for, say, a free muon, then we can expect that the dot product of these currents will have something to do with the electromagnetic interaction of these particles. Multiplying that by a propagator⁵⁵ (which contains the $g_{\mu\nu}$ tensor for the above dot product) allows for momentum transfer between the interacting particles and indeed gives the matrix element of the lowest-order approximation⁵⁶ to the scattering amplitude. That matrix element can be visualized by a Feynman diagram (see Fig. 6.5).

It is a property of the Standard Model (SM) that interactions between any two SM fermions can be described, in a leading approximation, by the current-current interaction as outlined above, although in the case of the weak interactions the left- and right-handed parts have to be treated separately because the weak interaction bosons couple to them differently (see Chapter 7). Since in the probability current there are no helicity-mixing terms like $\bar{u}_L^- \gamma^\mu u_R^+$, there are only two fundamental SM vertices, as illustrated in Fig. 6.6, where f stands for a fermion and the wiggly line of the exchanged particle represents one of the SM vector bosons: the photon, W^+ , W^- , Z^0 , or any of the eight gluons.

⁵⁵For an introduction to propagators, see Chapter 1 and, for example, [84] for greater detail.

⁵⁶In this context, ‘lowest order’ refers to a perturbative expansion in powers of the coupling constant—here $\alpha^2/4\pi$.

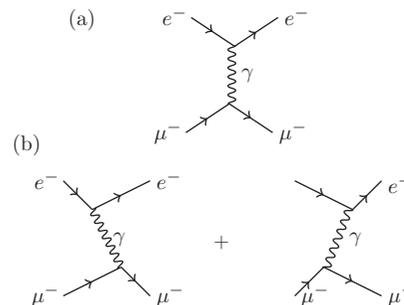


Fig. 6.5 (a) Feynman diagram for $e^- + \mu^- \rightarrow e^- + \mu^-$. This is in fact the sum of two diagrams. The vertical wiggly line representing the exchanged particle propagator is the sum of two scenarios depending on which particle was the emitter and which was the absorber of the exchanged particle (e.g. a photon or Z^0) as sketched in (b). Time goes from left to right.

The helicity is the same before and after the scattering (coupling to the exchanged particle). One says that the helicity is conserved, remaining unchanged by the interaction.

The annihilation or pair creation vertices are obtained from the scattering ones by *crossing symmetry* (see e.g. [13] or [84]). Keeping in mind that in Fig. 6.6, time is going from left to right, we ‘cross’ the incoming particle to the other side of the reaction equation by inverting its 4-momentum and swapping the helicity state so that it travels backwards in time, representing the outgoing antiparticle with the opposite helicity travelling forwards in time⁵⁷ (Fig. 6.6):

$$\Psi^+(x) = u^{(1)}(\mathbf{p})e^{-ip \cdot x} \rightarrow u^{(4)}(-\mathbf{p})e^{+ip \cdot x} = v^{(1)}(\mathbf{p})e^{+ip \cdot x}$$

⁵⁷The arrow on the fermion line indicates in which direction with respect to the time arrow the fermion specified by the label at the end of the line is travelling.

Crossing symmetry allows us to obtain the annihilation amplitude, for example for $e^+ + e^- \rightarrow \mu^+ + \mu^-$, from the scattering one, for $e^- + \mu^- \rightarrow e^- + \mu^-$, by ‘crossing’ the relevant particles to the other side of the reaction equation by changing in the scattering amplitude the corresponding 4-momenta $p \rightarrow -p$, the helicities, and the spinors $u \rightarrow v$, and—something that is beyond the formalism we are using—putting in ‘by hand’ the minus sign in front of the whole amplitude (QFT is needed

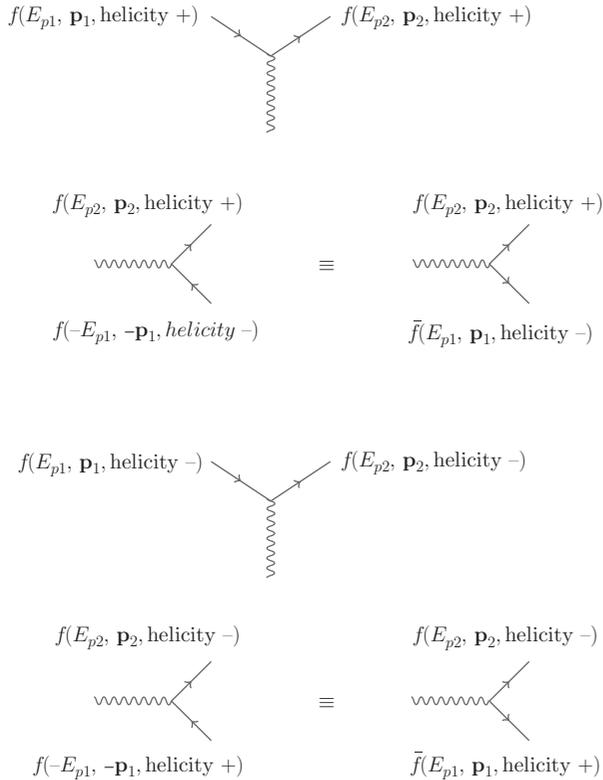


Fig. 6.6 Fundamental vertices of the Standard Model in the high-energy limit where masses can be neglected.

for that). This transformation also affects the description of the propagator. In the scattering amplitude, we have, in the example considered, 4-momenta p_1 and p_2 for the incoming and outgoing electrons, respectively, and therefore the 4-momentum of the exchanged boson is the difference $p_1 - p_2$, which, dotted with itself, gives $t \equiv (p_1 - p_2) \cdot (p_1 - p_2)$; for that reason, the scattering is called the t -channel process. With the change $p_2 \rightarrow -p_2$, $t \rightarrow s \equiv (p_1 + p_2) \cdot (p_1 + p_2)$, which is the square of the centre-of-mass energy; the resulting annihilation reaction is therefore called the s -channel process.⁵⁸

⁵⁸There is also a u -channel process; see e.g. [84].

6.4.4 P , T , and a comment on C

By construction, the Dirac equation is covariant with respect to Lorentz transformations and space inversion. It is also covariant with respect to time inversion. Space inversion $P : \mathbf{r} \rightarrow -\mathbf{r}$ and time inversion $T : t \rightarrow -t$ are discussed in this section. We make a comment about charge conjugation $C : \text{particle} \rightarrow \text{antiparticle}$ at the end of this section.

Consider two observers with reference frames \mathcal{O} and \mathcal{O}' . They are describing the same system or physical process, using their coordinates and wavefunctions $\Psi(x)$ and $\Psi'(x')$, respectively. The coordinates are related by a linear coordinate transformation $(x^\nu)' = a^\nu_\mu x^\mu$, where a^ν_μ could be the Lorentz transformation or a space or time inversion. The gamma matrices in their Dirac equations are also changed by that transformation, but, after a lot of algebra, it can be shown that both observers can neglect differences between their gamma matrices.⁵⁹ The covariance of the Dirac equation requires that the wavefunctions transform as⁶⁰

$$\Psi'(x') = \Psi'(ax) = S(a)\Psi(x) = S(a)\Psi(a^{-1}x')$$

where $S(a)$ is a matrix, $S^{-1}(a)$, exists and $S^{-1}(a) = S(a^{-1})$. One says that the coordinate transformation a induces a transformation $S(a)$ in the space of wavefunctions: $\Psi'(x') = S(a)\Psi(x)$.

Space inversion P

The space inversion transformation is

$$a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

We want to find $S(a)$ satisfying $\Psi'(x') = S(a)\Psi(x)$. Considering, for example, $\Psi^+(x) = u^{(1)}e^{-ip \cdot x}$ and following the expectation from classical

⁵⁹For proofs of all the results mentioned in this section, see e.g. [52].

⁶⁰It can also be shown that if the wavefunctions transform this way, then the Dirac equation is covariant with respect to this coordinate transformation.

physics that $\mathbf{r} \rightarrow -\mathbf{r}$ makes $\mathbf{r}' = -\mathbf{r}$, $\mathbf{p}' = -\mathbf{p}$ and $\boldsymbol{\sigma}' = \boldsymbol{\sigma}$ (the angular momentum axial vector), we obtain

$$\begin{aligned} (\Psi^+)'(x') &= \begin{pmatrix} \vartheta^{(1)} \\ \frac{\boldsymbol{\sigma}' \cdot \mathbf{p}'}{E_p + m} \vartheta^{(1)} \end{pmatrix} \exp[-i(E_p t - \mathbf{p}' \cdot \mathbf{x}')] \\ &= \begin{pmatrix} \vartheta^{(1)} \\ \frac{-\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m} \vartheta^{(1)} \end{pmatrix} \exp[-i(E_p t - \mathbf{p} \cdot \mathbf{x})] \\ &= \gamma^0 \Psi^+(x) \end{aligned}$$

giving $S = \gamma^0$, up to a fixed phase, which is set to 1 by convention. We could get that result immediately by looking at the transformation properties of Weyl spinors (see eqn 6.9) and the transformation from the Weyl to the Dirac representation (eqn 6.45).

Eigenstates of the parity operator $S = \gamma^0$ are states of defined intrinsic parity. Positive-energy states in the particle rest frame,

$$u^{(1)} e^{-im\tau}, \quad u^{(2)} e^{-im\tau}$$

are eigenstates of γ^0 with eigenvalue +1, i.e. they have intrinsic parity +1, but negative-energy states

$$v^{(1)} e^{+im\tau}, \quad v^{(2)} e^{+im\tau}$$

are eigenstates with eigenvalue -1 , thus having intrinsic parity -1 . We can state, then, that the intrinsic parity of a spin- $\frac{1}{2}$ particle (defined in its rest frame) is the negative of the intrinsic parity of its antiparticle. This applies to higher-spin fermions as well. Free-particle states with $\mathbf{p} \neq 0$ are not eigenstates of the parity operator γ^0 .

Time inversion T

For time inversion,

$$a = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The derivation of $S(a)$ here is more complicated than for space inversion. We consider only one aspect and then give the result. From classical physics, we expect that $t \rightarrow -t$ makes $\mathbf{r}' = \mathbf{r}$, $\mathbf{p}' = -\mathbf{p}$, and $t' = -t$, and therefore

$$\begin{aligned} \exp[-i(E_p t' - \mathbf{p}' \cdot \mathbf{x}')] &= \exp[-i(-E_p t + \mathbf{p} \cdot \mathbf{x})] \\ &= \{\exp[-i(E_p t - \mathbf{p} \cdot \mathbf{x})]\}^* \end{aligned}$$

so complex conjugation is involved. The full derivation gives $\Psi'(t') = S(a)\Psi(t) = i\gamma^1\gamma^3\Psi(t)^*$, again up to an arbitrary fixed phase, which, by convention, is taken to be 1.

It is useful to consider how $S(T) \equiv S(a)$ acts on a free-particle state, for example

$$\begin{aligned} S(T) \left(\begin{array}{c} \vartheta^{(1)} \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E_p + m} \vartheta^{(1)} \end{array} \right) \exp[-i(E_p t - \mathbf{p} \cdot \mathbf{x})] \\ = -i \left(\begin{array}{c} \vartheta^{(2)} \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}'}{E_p + m} \vartheta^{(2)} \end{array} \right) \exp[-i(E_p t' - \mathbf{p}' \cdot \mathbf{x}')] \end{aligned}$$

Note that $\vartheta^{(1)} \rightarrow \vartheta^{(2)}$, but the helicity does not flip as well because the direction of the momentum changes too, so a positive-helicity state remains a positive-helicity state in the time-reversed system.

$S(T)$ is anti-unitary, $S^2(T) = -1$, and this has interesting consequences. If we have interactions that are invariant with respect to time inversion, then $S(T)$ commutes with the Hamiltonian. If Ψ is an eigenstate of the Hamiltonian, then $S(T)\Psi$ is also an eigenstate of the Hamiltonian, with the same energy. But if $S(T)\Psi = \zeta\Psi$, where ζ is a phase, then

$$S(T)S(T)\Psi = S(T)\zeta\Psi = \zeta^*\zeta\Psi = \Psi$$

in contradiction to $S^2(T) = -1$. For that reason, $S(T)\Psi$ is a different state to Ψ ; there is at least a twofold degeneracy, known as Kramer's degeneracy. A spin- $\frac{1}{2}$ particle has a natural twofold degeneracy due to the two spin projections ($2j + 1$ states). A static magnetic field can lift that degeneracy by coupling to the particle's magnetic dipole moment, but a static magnetic field is not invariant with respect to time inversion, so nothing is wrong with having non-degenerate states in the magnetic field. The situation is different if the particle is put into a static electric field, which is invariant with respect to time inversion. If the particle has an electric dipole moment, the electric field will couple to it and would lift the $2j + 1$ degeneracy, shifting the energy levels up or down depending on the electric dipole orientation. Kramer's degeneracy forbids this because shifted energy states are required to be at least twofold-degenerate. So, unless there is an extra degree of freedom to guarantee this, electric dipole moments are forbidden by Kramer's degeneracy. A molecule of water has a large electric dipole moment—but in molecular or atomic systems, there are extra energy-degenerate (or nearly degenerate) states that allow this (see e.g. [126]).

In a series of beautiful experiments, pioneered by N. F. Ramsey [120, 121], using a beam of ultra-cold neutrons coming from a nuclear reactor at the Institut Laue-Langevin (ILL) in Grenoble, the absolute value of the neutron electron dipole moment d was measured to be smaller than $2.9 \times 10^{-26} e \text{ cm}$ at 90% confidence level.⁶¹

A comment on C

For every particle, there is a corresponding antiparticle (and vice versa). This symmetry of nature is called charge-conjugation symmetry.

⁶¹The SM does predict a very tiny violation of time-inversion symmetry, leading to the prediction of a very tiny d , much smaller than this limit. A number of extensions of the SM have been rejected because they predicted d to be larger than the current limit.

It has nothing to do with Lorentz invariance and is the same in all inertial frames. Charge conjugation leads to a unitary operator C in RQF (all states have positive energies in RQF). In RQM, one can construct an anti-unitary operator transforming a positive-energy state into a negative-energy state (spinor $u \rightarrow$ spinor v). Because it is anti-unitary, i.e. different from the common unitary charge-conjugation operator of RQF, we will not spend time deriving it, to avoid possible confusion.

We will, however, briefly outline basic properties of charge conjugation (see e.g. [118]). C changes the sign of the electric charge, magnetic moment, baryon number, and lepton number. Dynamical quantities like the energy, momentum, and helicity are left unchanged. Except for the weak interaction, all the interactions obey charge-conjugation symmetry. Eigenstates of charge conjugation are neutral particles like the photon, π^0 , η , and ρ^0 , with eigenvalue $+1$ or -1 . It is -1 for the photon and therefore $+1$ for a two-photon state, and so, as the π^0 has eigenvalue $+1$, the decay $\pi^0 \rightarrow \gamma\gamma$ is allowed. However, charge-conjugation symmetry forbids π^0 from decaying into three photons (eigenvalue -1). For particles built out of two fermions, $f\bar{f}$, the eigenvalue is $(-1)^{l+S}$, where S is the total spin of the $f\bar{f}$ state and l is the relative orbital angular momentum of the fermions. The e^+e^- bound state, positronium, decays to two photons if it is in the singlet state ($S = 0$, $l = 0$) and to three photons if it is in the triplet state ($S = 1$, $l = 0$).

In conclusion, we note that the weak interaction violates the symmetry of each of the three discrete transformations P , T , and C , as well as any superposition of any two of them. For all other interactions, each of the three discrete transformations is a symmetry of the interaction. But all known interactions, including the weak interaction, obey CPT symmetry.⁶² In RQF, it is impossible to construct Lorentz-invariant interactions that would violate CPT symmetry.

⁶²It should be noted that in RQF, CPT is an anti-unitary operator.

6.4.5 Electromagnetic interactions and the non-relativistic limit

So far, we have considered only the free-particle Dirac equation. The interaction of an electron (or indeed any electrically charged spin- $\frac{1}{2}$ structureless fermion) with a classical electromagnetic field is introduced through so-called ‘minimal coupling’, following the prescription by which the electromagnetic interactions are introduced in classical mechanics. The physics of this prescription will be discussed in Section 6.5.

The classical electromagnetic field is given by a 4-vector potential $A^\mu = (\phi, \mathbf{A})$ and its interaction is introduced into the Dirac equation by modifications of the derivatives:

$$\begin{aligned} i\frac{\partial}{\partial t} &\rightarrow i\frac{\partial}{\partial t} - q\phi \\ -i\nabla &\rightarrow -i\nabla - q\mathbf{A} \end{aligned}$$

where $q < 0$ is the electron charge. The Dirac equation then becomes

$$\left(i \frac{\partial}{\partial t} - q\phi \right) \Psi(\mathbf{r}, t) = [\boldsymbol{\alpha} \cdot (-i\nabla - q\mathbf{A}) + \beta m] \Psi(\mathbf{r}, t) \quad (6.54)$$

Solutions of this equation exist for some cases of the electromagnetic field (see e.g. [96]). We will find the non-relativistic limit of eqn 6.54 by applying an iterative method of approximation.

At low energies, the mass m of a particle is the main part of the energy, so we will factor that part out of the wavefunction:

$$\Psi(\mathbf{r}, t) = e^{-imt} \begin{pmatrix} \psi_U(\mathbf{r}, t) \\ \psi_L(\mathbf{r}, t) \end{pmatrix}$$

$\psi_U(\mathbf{r}, t)$ and $\psi_L(\mathbf{r}, t)$ are known as the upper and lower components of the Dirac spinor. They contain all ‘non-rest’ energy information relevant for the non-relativistic limit. Inserting this Ψ into eqn 6.54, we find

$$i \frac{\partial \psi_U}{\partial t} = \boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A}) \psi_L + q\phi \psi_U \quad (6.55)$$

$$i \frac{\partial \psi_L}{\partial t} = \boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A}) \psi_U + q\phi \psi_L - 2m\psi_L \quad (6.56)$$

Equation 6.56 may be rearranged to give

$$\psi_L = \frac{\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})}{2m} \psi_U - \frac{i \frac{\partial}{\partial t} - q\phi}{2m} \psi_L \quad (6.57)$$

Because m is relatively large, ψ_L is small compared with ψ_U . In the first-order approximation, we neglect the last term in eqn 6.57 and take

$$\psi_L \simeq \psi_{L1} = \frac{\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})}{2m} \psi_U$$

Inserting this into eqn 6.55, we obtain the first-order approximation ψ_{U1} for ψ_U :

$$i \frac{\partial \psi_{U1}}{\partial t} = q\phi \psi_{U1} + \frac{[\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})][\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})]}{2m} \psi_{U1}$$

⁶³Namely,

$$\begin{aligned} (\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{A}) &= \mathbf{A} \cdot \mathbf{A} + i\boldsymbol{\sigma} \cdot (\times \mathbf{A}) \\ (\nabla \times \mathbf{A} + \mathbf{A} \times \nabla)f &= f(\nabla \times \mathbf{A}) = f\mathbf{B} \end{aligned}$$

Using vector identities,⁶³ this equation can be written as

$$i \frac{\partial \psi_{U1}}{\partial t} = q\phi \psi_{U1} + \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} \psi_{U1} - \frac{q}{2m} \boldsymbol{\sigma} \cdot (\nabla \times \mathbf{A} + \mathbf{A} \times \nabla) \psi_{U1}$$

and finally as

$$i \frac{\partial \psi_{U1}}{\partial t} = H_P \psi_{U1}$$

where

$$H_P \equiv q\phi + \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} - \frac{q}{2m} \boldsymbol{\sigma} \cdot \mathbf{B}$$

is the Pauli Hamiltonian for the Schrödinger equation and \mathbf{B} is the magnetic field. Since the Hamiltonian of the interaction of a magnetic dipole moment $\boldsymbol{\mu}$ with an external magnetic field \mathbf{B} is $-\boldsymbol{\mu} \cdot \mathbf{B}$, we can identify $(q/2m)\boldsymbol{\sigma}$ with the electron magnetic dipole moment $\boldsymbol{\mu}$. On the other hand, the electron magnetic moment is related to the electron spin by $\boldsymbol{\mu} = -g\frac{1}{2}\boldsymbol{\sigma}\mu_B$, where g is the proportionality factor and $\mu_B = |q|/2m$ is the Bohr magneton (q is the electron charge). From this, we find that $g = 2$, which was a triumph for Dirac and his equation.

The value of g was known from atomic physics measurements, but until the advent of the Dirac equation, there was no explanation why the experimental value was as it was. In fact, the value of g for the electron, as well as for the muon, is not exactly 2. The small difference is explained by RQF. The value of $(g - 2)/2$ has been measured with fantastic precision: 2.8×10^{-13} for the electron [85] and 6×10^{-10} for the muon [49]. As for the precision measurements of the neutron electric dipole moment, in these cases also the precision of the measurements and theoretical calculations give constraints on extensions of the SM.

It should be noted that the magnetic dipole moment of the proton is quite different from that predicted by the Dirac equation. The value of g is about 5.6 instead of about 2 (using the proper magneton for the proton). For the neutron, it is even more surprising—instead of 0 because the electric charge is 0, g is about -3.8 . These significant deviations from the predictions of the Dirac equation, which are applicable for *point-like* particles, were the first indications that protons and neutrons are not point-like particles. As we now know, they have a complicated substructures of quarks and gluons.

We can carry on with the iterative procedure and get relativistic corrections beyond the Pauli Hamiltonian. The next one is obtained by substituting ψ_{L1} for ψ_L in the last term of eqn 6.57, thus giving the second-order approximation:

$$\psi_L \simeq \psi_{L2} = \frac{\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})}{2m} \psi_U - \frac{\left(i\frac{\partial}{\partial t} - q\phi\right) [\boldsymbol{\sigma} \cdot (\mathbf{p} - q\mathbf{A})]}{2m} \psi_U$$

Unfortunately, inserting this into eqn 6.55 gives a Hamiltonian that is not Hermitian and the electron acquires an imaginary electric dipole moment. Formal problems of this type took many years to solve after Dirac published his equation. Eventually, a consistent second-order Hermitian Hamiltonian was found:

$$\begin{aligned} H = & \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} - \frac{|\mathbf{p}|^4}{8m^3} \\ & + q\phi - \frac{q}{2m} \boldsymbol{\sigma} \cdot \mathbf{B} \\ & - \frac{iq}{8m^2} \mathbf{p} \cdot \mathbf{E} \\ & - \left[\frac{iq}{8m^2} \boldsymbol{\sigma} \cdot (\nabla \times \mathbf{E}) + \frac{q}{4m^2} \boldsymbol{\sigma} \cdot (\mathbf{E} \times \mathbf{p}) \right] \end{aligned}$$

The first line in this expression for the Hamiltonian represents the kinetic term with its first relativistic correction, the third the Darwin term, and the fourth the spin-orbit interaction (where \mathbf{E} is the electric field). For $\mathbf{E} = -\nabla\phi$ and a spherically symmetric potential ϕ , $\boldsymbol{\sigma} \cdot (\nabla \times \mathbf{E}) = 0$ and the spin-orbit term takes the familiar form

$$\begin{aligned} H_{\text{SO}} &= -\frac{q}{4m^2} \boldsymbol{\sigma} \cdot (\mathbf{E} \times \mathbf{p}) \\ &= \frac{q}{4m^2} \boldsymbol{\sigma} \cdot \frac{\partial\phi}{\partial r} \frac{\mathbf{r}}{r} \times \mathbf{p} \\ &= \frac{q}{4m^2} \frac{1}{r} \frac{\partial\phi}{\partial r} \boldsymbol{\sigma} \cdot \mathbf{l} \end{aligned}$$

where \mathbf{l} is the orbital angular momentum of the electron. It should be noted that Thomas precession is automatically included, as it should be, in the relativistic formalism.

Historically, the first formally successful derivation of the non-relativistic limit of the Dirac equation that included interactions with the (classical) electromagnetic field was obtained via the FW transformation. It is interesting that one has to introduce the interaction first into the Dirac equation in the Weyl or Dirac representation (or any other representation related to them by a transformation not depending on the momentum) and only then make the FW transformation. One might think that doing the FW transformation for the free-particle Dirac equation first, thus decoupling the lower and upper components of the Dirac spinor, and then introducing the interactions via ‘minimal coupling’ would work—but it does not.

6.5 Gauge symmetry

By considering transformations of space-time, one gets a relativistic description of free particles. In order to describe their interactions, one needs to consider transformations, gauge transformations, in another space, an internal space. Symmetries of the gauge transformations in that internal space are at the heart of the SM. Before discussing gauge symmetries, we need to revise the three essential ingredients of the formalism.

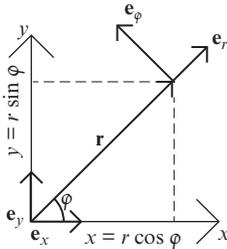


Fig. 6.7 Cartesian and spherical coordinate systems.

6.5.1 Covariant derivative

Consider two coordinate systems as sketched in Fig. 6.7. The spherical basis vectors are related to the Cartesian ones in the following way:

$$\begin{aligned} \mathbf{e}_r &= \frac{\partial x}{\partial r} \mathbf{e}_x + \frac{\partial y}{\partial r} \mathbf{e}_y = \cos \varphi \mathbf{e}_x + \sin \varphi \mathbf{e}_y, & |\mathbf{e}_r| &= 1 \\ \mathbf{e}_\varphi &= \frac{\partial x}{\partial \varphi} \mathbf{e}_x + \frac{\partial y}{\partial \varphi} \mathbf{e}_y = -r \sin \varphi \mathbf{e}_x + r \cos \varphi \mathbf{e}_y, & |\mathbf{e}_\varphi| &= r \end{aligned}$$

Next introduce a constant vector field \mathbf{a} of unit length, as sketched in Fig. 6.8, using both coordinate systems. In the Cartesian basis,

$$\mathbf{a} = 1 \mathbf{e}_x + 0 \mathbf{e}_y = a^x \mathbf{e}_x + a^y \mathbf{e}_y$$

and in the spherical basis,

$$\mathbf{a} = \cos \varphi \mathbf{e}_r - \frac{1}{r} \sin \varphi \mathbf{e}_\varphi = a^r \mathbf{e}_r + a^\varphi \mathbf{e}_\varphi$$

Although

$$\frac{\partial a^x}{\partial x} = \frac{\partial a^x}{\partial y} = \frac{\partial a^y}{\partial x} = \frac{\partial a^y}{\partial y} = 0$$

in the spherical basis,

$$\frac{\partial a^r}{\partial \varphi} = -\sin \varphi \neq 0, \quad \frac{\partial a^\varphi}{\partial \varphi} = -\frac{1}{r} \cos \varphi \neq 0$$

which looks wrong because the field \mathbf{a} is constant—nothing is changing from point to point. It is wrong because the differentiation has not taken into account that the spherical basis vectors change from one space point to another. If we take this into account, differentiating not only coordinates but also basis vectors, everything is fine. For example,

$$\begin{aligned} \frac{\partial \mathbf{a}}{\partial \varphi} &= \frac{\partial}{\partial \varphi} \left(\cos \varphi \mathbf{e}_r - \frac{1}{r} \sin \varphi \mathbf{e}_\varphi \right) \\ &= \frac{\partial}{\partial \varphi} (\cos \varphi) \mathbf{e}_r + \cos \varphi \frac{\partial}{\partial \varphi} (\mathbf{e}_r) \\ &\quad + \frac{\partial}{\partial \varphi} \left(-\frac{1}{r} \sin \varphi \right) \mathbf{e}_\varphi - \frac{1}{r} \sin \varphi \frac{\partial}{\partial \varphi} (\mathbf{e}_\varphi) \\ &= -\sin \varphi \mathbf{e}_r + \cos \varphi \left(\frac{1}{r} \mathbf{e}_\varphi \right) - \frac{1}{r} \cos \varphi \mathbf{e}_\varphi - \frac{1}{r} \sin \varphi (-r \mathbf{e}_r) \\ &= 0 \end{aligned}$$

So, in general, for a vector \mathbf{V} ,

$$\frac{\partial \mathbf{V}}{\partial x^\beta} = \frac{\partial V^\alpha}{\partial x^\beta} \mathbf{e}_\alpha + V^\alpha \frac{\partial \mathbf{e}_\alpha}{\partial x^\beta}$$

The last derivative, $\partial \mathbf{e}_\alpha / \partial x^\beta$, is a vector and can be described as a linear combination of the basis vectors \mathbf{e}_α :

$$\frac{\partial \mathbf{e}_\alpha}{\partial x^\beta} = \Gamma_{\alpha\beta}^\mu \mathbf{e}_\mu$$

The geometrical object $\Gamma_{\alpha\beta}^\mu$ is called a connection. Changing the names of the indices in the above equation, $\mu \rightarrow \alpha$ and $\alpha \rightarrow \mu$, we can write

$$\frac{\partial \mathbf{V}}{\partial x^\beta} = \left(\frac{\partial V^\alpha}{\partial x^\beta} + V^\mu \Gamma_{\mu\beta}^\alpha \right) \mathbf{e}_\alpha$$

This is the covariant derivative, which takes care of the changing coordinates as well as the basis vectors.

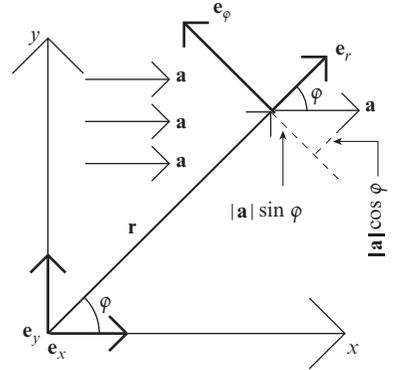


Fig. 6.8 A constant vector field \mathbf{a} .

6.5.2 Gauge invariance in electromagnetism

In classical electromagnetism, because $\nabla \cdot \mathbf{B} = 0$, we can introduce a vector potential \mathbf{A} , such that the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$. Then

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = 0 \quad \text{can be written as} \quad \nabla \times \left(\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = 0$$

allowing the introduction of a scalar potential ϕ , such that

$$\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} = -\nabla \phi$$

which, after rearrangement, gives the electric field as

$$\mathbf{E} = -\nabla \phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}$$

If we now take an arbitrary but differentiable scalar function $\lambda(\mathbf{r}, t)$ and make the transformation

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla \lambda \quad (6.58)$$

then \mathbf{B} remains unchanged. If, simultaneously with this transformation, we make the change

$$\phi \rightarrow \phi' = \phi - \frac{1}{c} \frac{\partial \lambda}{\partial t} \quad (6.59)$$

then \mathbf{E} will also stay unchanged. The transformations in eqns 6.58 and 6.59 are called gauge transformations⁶⁴ and the fact that electric and magnetic fields stay the same is called gauge invariance.

⁶⁴In manifestly Lorentz-covariant form (and with $c = 1$),

$$A^\mu \rightarrow A'^\mu = A^\mu - \partial^\mu \lambda$$

6.5.3 The Aharonov–Bohm effect

Following Feynman et al. [75], we will examine two-slit electron diffraction. To study the Aharonov–Bohm effect, we use the set-up sketched in Fig. 6.9. With no current in the solenoid, electrons are diffracted by the slits and form an interference pattern on the screen. As soon as current is flowing through the solenoid, the diffraction pattern changes to another one. The probability amplitude ψ_1 for an electron to follow path 1 and the amplitude ψ_2 for the path 2 are modified in the following way:

$$\psi_1 = \psi_{01} \exp\left(-\frac{iqS_1}{\hbar}\right), \quad \psi_2 = \psi_{02} \exp\left(-\frac{iqS_2}{\hbar}\right)$$

where ψ_{01} and ψ_{02} are the amplitudes without the current, q is the electron charge, and S_1 and S_2 are extra phases due to the presence of the current in the solenoid, which are given by

$$S_1 = \frac{q}{\hbar} \int_{\text{path1}} \mathbf{A} \cdot d\mathbf{r}, \quad S_2 = \frac{q}{\hbar} \int_{\text{path2}} \mathbf{A} \cdot d\mathbf{r}$$

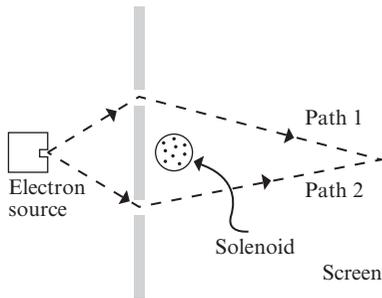


Fig. 6.9 Diagram of a two-slit electron diffraction experiment demonstrating the Aharonov–Bohm effect.

The modification of the phase on the screen is then

$$\begin{aligned} \frac{q}{\hbar}(S_1 - S_2) &= \frac{q}{\hbar} \left(\int_{\text{path1}} \mathbf{A} \cdot d\mathbf{r} - \frac{q}{\hbar} \int_{\text{path2}} \mathbf{A} \cdot d\mathbf{r} \right) \\ &= \frac{q}{\hbar} \oint_{\text{closed path}} \mathbf{A} \cdot d\mathbf{r} \end{aligned}$$

which, by Stokes' theorem, is proportional to the magnetic flux in the solenoid.

For an infinitely long thin (and therefore not obstructing the slits) solenoid, there is no magnetic field \mathbf{B} outside the solenoid. The experimentally observed modification of the interference pattern [60] is therefore due to the vector potential \mathbf{A} , present outside the solenoid, with a magnitude inversely proportional to the distance from the solenoid axis. Classically, the magnetic field \mathbf{B} and the vector potential \mathbf{A} are equivalent, in the sense that one can use either. This is not true at the quantum level, where \mathbf{A} is apparently more fundamental. It should be noted that the same applies to the scalar potential ϕ and the electric field \mathbf{E} , where instead of space dimensions one considers time.

The final conclusion that we take from the Aharonov–Bohm effect is the observation that the vector potential \mathbf{A} is related to the space-dependent phase of the probability amplitude and the scalar potential ϕ is related to the time-dependent phase. So, from the phase of the probability amplitude, we can get the 4-vector potential, and the other way around.

6.5.4 Interactions from gauge symmetry

The electromagnetic interactions have been introduced into the Dirac equation following the classical ‘minimal coupling’ procedure, which required modification of the derivatives:

$$\frac{\partial}{\partial t} \rightarrow D^0 \equiv \frac{\partial}{\partial t} + iq\phi, \quad \nabla \rightarrow \mathbf{D} \equiv \nabla - iq\mathbf{A}$$

which can be written in manifestly Lorentz-covariant form as

$$D^\mu \equiv \partial^\mu + iqA^\mu \quad (6.60)$$

The derivative D^μ is the covariant derivative and we can start thinking about the scalar potential ϕ and the vector potential \mathbf{A} as connections in a space to be defined. The outcome is the Dirac equation for the electron interacting with a classical electromagnetic field represented by the 4-vector potential (ϕ, \mathbf{A}) :

$$\left(i \frac{\partial}{\partial t} - q\phi \right) \Psi(\mathbf{r}, t) = [\boldsymbol{\alpha} \cdot (-i\nabla - q\mathbf{A}) + \beta m] \Psi(\mathbf{r}, t) \quad (6.61)$$

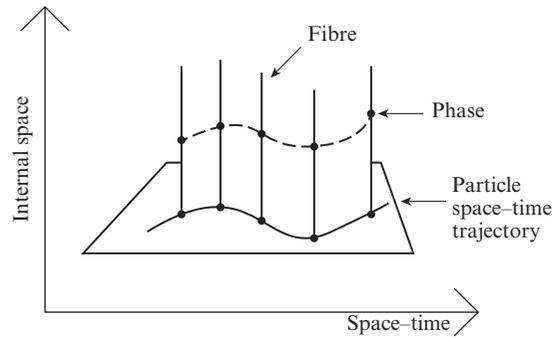


Fig. 6.10 A particle travelling in space-time and its internal space. Adapted from [111].

We know that the potentials are not unique and we can perform the gauge transformations 6.58 and 6.59 without affecting Maxwell’s equations and their solutions. Would the same apply to the Dirac equation 6.61? The answer is negative. A solution of eqn 6.61 will differ from that obtained by solving eqn 6.61 after the transformations 6.58 and 6.59. But if, simultaneously with 6.58 and 6.59, we also make the transformation

$$\Psi(\mathbf{r}, t) \rightarrow \Psi'(\mathbf{r}, t) = \exp[iq\lambda(\mathbf{r}, t)]\Psi(\mathbf{r}, t) \quad (6.62)$$

then the Dirac equation 6.61 will be covariant with respect to the three combined gauge transformations 6.58, 6.59, and 6.62, and the solution will describe the same physics as that of eqn 6.61 before the transformations. We see that changes to the 4-vector potential affect the space-time-dependent phase of the wavefunction; it needs to be changed accordingly as well. As in the Aharonov–Bohm effect, they are connected.

Now comes the crucial step. We reverse the flow of arguments. We first demand that we want interactions that are invariant with respect to the transformation 6.62. For that to happen, we need to modify the derivatives of the free-particle Dirac equation in such a way that the symmetry is obeyed. This requires a move from space-time derivatives to covariant derivatives and the introduction of the 4-vector potential (ϕ, \mathbf{A}) . So, demanding gauge symmetry with respect to the gauge transformation 6.62, we get the classical electromagnetic field with which our particle is interacting. If, in addition, in all formulae, for example the one for the probability current, we replace derivatives by covariant derivatives, they will also be form-invariant.

The last step is to introduce the internal space with its basis vectors on which the connection (the 4-vector potential) operates and extend the formalism to other interactions beyond electromagnetism.⁶⁵

We can imagine that a particle, moving in space-time, is carrying its internal space with it as sketched in Fig. 6.10. In mathematical language, the internal space is called a fibre and the whole structure, which locally is a product of the fibre and space-time, is called a fibre bundle. In the case of electromagnetism, the fibre is a circle, as sketched in Fig. 6.11.

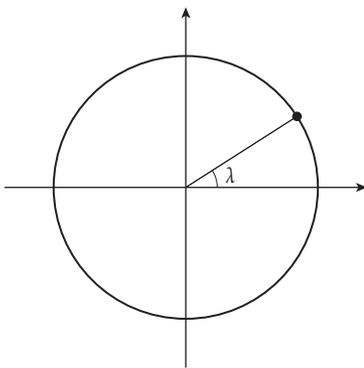


Fig. 6.11 The internal space of electromagnetism. Adapted from [111].

⁶⁵The formalism will only be outlined here. For more details, see [111].

Each point on the circle, parameterized by a real function $\lambda(\mathbf{r}, t)$, corresponds to a complex number with modulus 1, i.e. a one-dimensional unitary matrix $\exp[iq\lambda(\mathbf{r}, t)]$. The set of all such matrices forms a group called $U(1)$, with ‘1’ for one-dimensionality. When the particle moves from one space–time point to another, the phase of its wavefunction is modified (eqn 6.62) and the 4-vector potential (ϕ, \mathbf{A}) is modified as well (eqns 6.59 and 6.58). A basis in the space $U(1)$ consists of one particular matrix, for example the unit matrix (in this case the real number 1), which corresponds to $\lambda = 0$.

When the particle moves in space–time, $\lambda(\mathbf{r}, t)$ changes corresponding to the changing basis.⁶⁶ The change of basis is represented by the connection, which in turn gives us the change in the 4-vector potential (ϕ, \mathbf{A}) , eqns 6.59 and 6.58. So our ‘differential’ equation is: the connection equals the 4-vector potential.

We are ready now to extend the formalism to include other interactions. What will happen if in eqn 6.62, instead of the real function $\lambda(\mathbf{r}, t)$, we insert a matrix $\mathcal{M}(\mathbf{r}, t)$?⁶⁷ Suppose that \mathcal{M} belongs to $SU(2)$, the group of 2×2 complex unitary matrices with unit determinant (as indicated by the letter S \equiv ‘special’). $SU(2)$ has 3 basis vectors, which can be the Pauli matrices σ_x , σ_y , and σ_z . Each element of the group can then be defined by 3 real numbers λ^1 , λ^2 , and λ^3 , and can be represented as a point within a sphere of radius 2π in 3 dimensions. So our particle carries with itself such a sphere, its internal space, and the interaction that we get in this way will be the weak interaction. Generalizing eqn 6.62, we require gauge symmetry with respect to the following transformation:

$$\Psi(\mathbf{r}, t) \rightarrow \Psi'(\mathbf{r}, t) = \exp \left[iq \sum_{k=1}^3 \lambda^k(\mathbf{r}, t) \sigma_k \right] \Psi(\mathbf{r}, t) \quad (6.63)$$

The operator acting on the wavefunction Ψ is now a series of 2×2 matrices and therefore our wavefunction Ψ gets one extra dimension and is represented by two components (for historical reasons called projections of the isotopic spin) related to the $SU(2)$ internal space. In order to get the 4-vector potential of the weak interaction, we consider the infinitesimal change in the wavefunction when the space–time point is changed from (\mathbf{r}, t) to $(\mathbf{r} + d\mathbf{r}, t + dt)$. All components of the wavefunction are affected, but to get the 4-vector potential we select only the connection part of the covariant derivative, i.e. the change in the basis vectors in the internal space. That change is represented by the changes in λ^1 , λ^2 , and λ^3 , giving, finally,

$$A^\mu = \sum_{k=1}^3 [\partial^\mu \lambda^k(\mathbf{r}, t)] \sigma_k \quad (6.64)$$

Therefore, if we want to introduce the weak interactions into the free-particle Dirac equation, we must change space–time derivatives to

⁶⁶With respect to four coordinates, requiring four derivatives: one time and three space directions.

⁶⁷Noting that

$$\exp(\mathcal{M}) \equiv 1 + \mathcal{M} + \frac{1}{2}\mathcal{M}^2 + \dots$$

⁶⁸The 4-vector potential is a 2×2 matrix in the isotopic spin space generated by the Pauli matrices.

covariant derivatives, including the connection, i.e. the 4-vector potential given by eqn 6.64.⁶⁸

The strong interaction is introduced in a similar way. The internal space is now $SU(3)$, a group of unitary complex matrices with unit determinant. There are 8 basis vectors in that space, and therefore 8 real numbers identify each matrix belonging to the group. The Pauli matrices generating the weak interactions are replaced by these 8 basis matrices of the $SU(3)$ group and the wavefunction gains extra degrees of freedom, becoming a 3-dimensional column vector in the colour space of the strong interaction. The corresponding 4-vector potential becomes a 3×3 matrix. Changing derivatives to covariant derivatives causes the particle to interact with the classical colour field.

After quantization of the classical weak and strong fields, we have three spin-1 bosons for the weak interaction and 8 spin-1 gluons for the strong interaction. All of them are massless at this stage. To get massive physical Z^0 , W^+ and W^- bosons, we need an extra mechanism, like the Higgs mechanism that will be discussed in Chapter 12. Elements of $U(1)$ commute among themselves, but those of $SU(2)$ or $SU(3)$ do not. The physical consequences are that photons do not interact with each other but weak-interaction bosons and strong-interaction gluons do.

For further reading, [13] is particularly recommended.

Chapter summary

- Special relativity, Lorentz transformation, invariance, and covariance
- Klein–Gordon and Dirac equations, covariant derivative
- Introduction of electromagnetic interactions, spin–orbit interaction

Further reading

- Halzen, F. and Martin, A. D. (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley. This is an excellent book on RQM and particle physics, very well aligned with this book.
- Aitchison, I. J. R. and Hey, A. J. G. (2013). *Gauge Theories in Particle Physics*, Volumes 1 and 2 (4th edn). CRC Press. This is an excellent introduction to field theory in particle physics, starting from quantum mechanics.
- Barut, A. O. and Raączka (1986). *Theory of Group Representations and Applications* (2nd revised edn). World Scientific. This classical text on groups and their representations is recommended for readers who are theory-oriented.
- Steane, A. M. (2012). *Relativity Made Relatively Easy*. Oxford University Press. This book is recommended for its coverage of relativity, spinors, and the Dirac equations; it is very well aligned with this book.
- Berestetskii, W. B., Lifshitz, E. M., and Pitaevskii, L. P. (1971). *Relativistic Quantum Theory, Part I*. Pergamon Press. From the Landau course, this is a classical text, and still very good.
- Dirac, P. A. M. (1958). *Principles of Quantum Mechanics* (4th edn). Oxford University Press. See Chapter IX.

Exercises

- (6.1) Using the definitions given in Section 6.3, derive the continuity equation 6.21 and show that it may also be written in the covariant form $\partial_\mu j^\mu = 0$.
- (6.2) Follow Dirac's derivation of 'a relativistic equation for the electron'. Start from $(p^2 - m^2)\psi = 0$, where $p^2 = p_0^2 - |\mathbf{p}|^2$ and $c = 1$. Factorize the 4-momentum operator as

$$(p_0 + \sqrt{\mathbf{p}^2 + m^2})(p_0 - \sqrt{\mathbf{p}^2 + m^2})$$

and apply only the second bracket to give

$$(p_0 - \sqrt{\mathbf{p}^2 + m^2})\psi = 0$$

The difficulty is that quantum mechanics gives operator expressions for p_0 and p_i , which in the Schrödinger representation are first-order derivatives in time and space, respectively. The spatial derivatives are now under a square root. Dirac solved the problem by writing

$$\sqrt{\mathbf{p}^2 + m^2} = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta$$

where

$$p_r = -i\hbar \frac{\partial}{\partial x_r} \quad (r = 1, 2, 3), \quad p_0 = i\hbar \frac{\partial}{\partial t}$$

Show that to satisfy $E^2 = p^2 + m^2$, it is necessary that

$$\begin{aligned} \alpha_i \alpha_j + \alpha_j \alpha_i &= 2\delta_{ij} \\ \alpha_i \beta + \beta \alpha_i &= 0 \end{aligned}$$

($i, j = 1, 2, 3$), with $\beta^2 = 1$.

Dirac showed that the simplest realization of these constraints is given by 4-dimensional matrices. It should be noted that at this stage of the historical development, the wavefunction ψ had unknown Lorentz transformation properties; they had to be derived next.

- (6.3) The covariant Dirac matrices are defined by

$$\gamma^0 = \beta, \quad \gamma^i = \beta \alpha_i$$

so that $\gamma \cdot p$ is a 4-vector scalar product—often written as \not{p} , with the Dirac equation then being $(\not{p} - m)\psi = 0$. Using the definitions of the Dirac

matrices given in Section 6.4, follow the text to derive the defining relation

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu} \quad (6.65)$$

- (6.4) Using the information given at the end of Section 6.4.2 on helicity conservation in π^\pm decays, estimate the relative decay rates for $e\nu_e$ and $\mu\nu_\mu$ modes. (The two-body phase factor is discussed in Chapter 2.)

- (6.5) Write down the Dirac equation for free electrons. Derive the adjoint Dirac equation and an expression for the probability current 4-vector.

The free-electron positive-energy solutions of the Dirac equation in the standard representation are

$$\psi_s(\mathbf{r}, t) = N \begin{pmatrix} \chi^s \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E + m} \chi^s \end{pmatrix} e^{i(\mathbf{p} \cdot \mathbf{r} - Et)/\hbar}$$

where $s = 1, 2$, $\chi^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $\chi^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Find a normalization factor N and explain why these wavefunctions are not normalized to one particle per unit volume.

The operator for spin projection on the x axis is

$$\Sigma_1 = \frac{1}{2} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_1 \end{pmatrix}$$

Find its eigenvalues and eigenvectors, and show that in the non-relativistic limit one can form linear combinations of $\psi_1(\mathbf{r}, t)$ and $\psi_2(\mathbf{r}, t)$ that are eigenvectors of Σ_1 but at relativistic energies this is impossible.

- (6.6) Given that $\Psi(x) = u(p)e^{-ip \cdot x}$ is a solution of the Dirac equation, where $\bar{u}(p)u(p) = 2m$, derive expressions for $u(p)$ with $p = (E, 0, 0, p_z)$. Here, p is the 4-momentum, $p \cdot p = m^2$, and x is the 4-displacement.

What are the $u(p)$ spinors in the ultrarelativistic limit? What are their helicities? Why is the Dirac representation more suitable in the low-energy limit?

- (6.7) A free-particle solution of the Dirac equation is $\Psi(x) = C(p)e^{-ip \cdot x}$, where $p = (E, 0, 0, p_3)$. Find the normalized $E > 0$ spinors $C_+(p)$ and $C_-(p)$

for positive- and negative-helicity states in the standard Dirac representation and using covariant normalization.

The operator for spin projection on the x axis is

$$\Sigma_1 = \frac{1}{2} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_1 \end{pmatrix}$$

Show that neither $C_+(p)$ nor $C_-(p)$ is an eigenstate of Σ_1 .

The expectation value of Σ_1 is

$$\langle \Sigma_1 \rangle = \frac{1}{2E} C^\dagger(p) \Sigma_1 C(p)$$

where $C = C_+(p) \cos \alpha + C_-(p) \sin \alpha$, with α a real constant, and normalization is to unit volume. Calculate $\langle \Sigma_1 \rangle$ and interpret the result in the non-relativistic and ultrarelativistic limits for $\alpha = \pm \frac{1}{4} \pi$.

(6.8) Give the expression for the total energy operator H , including the electromagnetic potential (A^0, \mathbf{A}) , for a particle of mass m .

Introducing the upper and lower bispinor components (φ, χ) of ψ and writing $H = m + H_n$, derive an expression for $H_n \psi$. Using $H_n \psi$, show that, under certain conditions to be specified, $\chi \sim (\text{velocity}/c) \times \varphi$, for a stationary state $H\psi = E\psi$.

Under these conditions, show how the Dirac equation reduces to the non-relativistic Schrödinger–Pauli equation for a spin- $\frac{1}{2}$ particle with energy $E_n = E - m$. Identify the term giving the interaction of the particle’s spin with the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ and comment on its magnitude.

[The identity $(\boldsymbol{\sigma} \cdot \mathbf{a})(\boldsymbol{\sigma} \cdot \mathbf{b}) = \mathbf{a} \cdot \mathbf{b} + i\boldsymbol{\sigma} \cdot \mathbf{a} \times \mathbf{b}$ may be assumed.]

Weak interactions

7

This chapter introduces the theory of weak interactions and electroweak unification. The experimental evidence for this theory will be reviewed in Chapter 8. We give a description of electroweak interactions with emphasis on the key physics ingredients. More mathematical treatments (graduate-level textbooks) are listed in Further Reading at the end of the chapter. The Higgs mechanism and supporting experimental evidence will be covered in Chapter 12.

The weak force is responsible for the β decay of nuclei, for the decay of the lightest hadrons such as the pion, kaon, and neutron, and for neutrino interactions. The weak interaction violates spatial parity conservation and facilitates flavour-changing interactions. The quanta of the weak force are the massive W^\pm and Z^0 . Weak interactions mediated by the W^\pm are termed ‘charged-current’ (CC) and those by the Z^0 ‘neutral-current’ (NC). In this chapter, the evidence for the ‘universality’ of the weak interaction is outlined and its importance explained. Figure 7.1 shows neutron β decay at the quark level; it is a CC decay. At this energy scale (~ 1 GeV), the W will hardly propagate and it is safe to approximate the propagator $g_w^2/(M_W^2 - q^2)$ by a constant, which we will see is closely related to the Fermi constant G_F (the concept of a force propagator is outlined at the start of Chapter 9).

The proposal by Glashow, Weinberg, and Salam (independently) that the weak and electromagnetic interactions should be unified was a key step for particle physics. At first sight, this seems difficult to achieve because the photon is massless but the weak bosons are not—the Higgs mechanism for mass generation is crucial for this. The discovery of ‘weak neutral current’ interactions in neutrino scattering was the first indication that these ideas might be correct. The discovery in 1983 by the UA1 and UA2 experiments at the $Spp\bar{p}S$ collider at CERN of the W^\pm and Z^0 particles with masses close to the predicted values was a seminal moment in the development of the Standard Model.

Other topics covered in this chapter are

- the generalization of the idea of a current–current interaction;
- left- and right-handed particles;
- the CKM matrix, which connects the quark mass eigenstates with the weak eigenstates;
- the GIM mechanism, which explains how flavour-changing weak decays are suppressed despite the universality of the weak force.

7.1 Fermi theory	182
7.2 Weak interactions of leptons	183
7.3 Weak interactions including quarks	189
7.4 Introduction to electroweak unification	193
Chapter summary	202
Further reading	203
Exercises	203

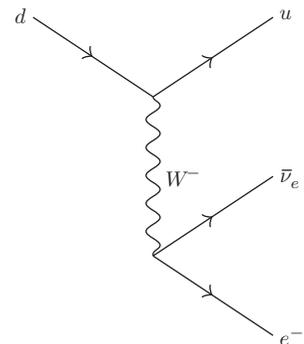


Fig. 7.1 Neutron β decay at the quark level, a CC weak decay. Note that the spectator ud pair in the n and p are omitted.

Much of the early work on weak interactions focused on understanding the seeming complexity of particle decays. Experiments made possible by the high energies of the LEP and LHC colliders have enabled the richness and underlying symmetry of electroweak interactions to be understood at a more profound level. After a brief review of the Fermi theory of β decay, the chapter is divided into three main sections dealing with weak interactions of leptons, weak interactions including quarks, and electroweak unification, respectively.

7.1 Fermi theory

The original theory of nuclear β decay, by Fermi, uses his ‘Golden Rule’ of quantum mechanics to calculate rates. For example, the decay rate of the neutron ($n \rightarrow pe^- \nu_e$) is given by

$$w = \frac{2\pi}{\hbar} |M|^2 \frac{dN}{dE} \quad (7.1)$$

where M is the matrix element given by

$$M = \int \psi_f O \psi_i dV = \int \psi_p^* \psi_e^* \psi_\nu^* O \psi_n dV \quad (7.2)$$

M involves the initial- and final-state wavefunctions ψ_i and ψ_f , respectively, and O is an operator describing the interaction, which was assumed to be local, with a strength given by G_F . This amounts to ignoring any propagator of the weak force. Given the very large mass of the W^\pm boson (~ 80 GeV), for decays of leptons and most hadrons (masses up to a few GeV), the propagator $g_w^2/(M_W^2 - q^2)$ can be approximated by the constant $G_F \sim g_w^2/M_W^2$. The wavefunctions of the electron and the neutrino were assumed to be constant over the volume of the nucleus and were moved outside the integral; the volume over which they are normalized cancels with the volume used to calculate the density of states dN/dE .¹ Another simplification was to use Schrödinger wavefunctions ψ rather than the Dirac spinors that are needed for a correct relativistic description of spin- $\frac{1}{2}$ particles.

Some nuclear decays involve larger changes in nuclear spins than can be accommodated by the spin 0 or 1 change from the electron–neutrino pair. These decays tend to proceed more slowly. This leads to a jargon of ‘forbidden decays’—for which the decay could occur with a change in the nuclear orbital angular momentum but more slowly than the ‘allowed decays’. The terminology is a bit unfortunate, but it does introduce a very important idea. If a process occurs at a slower rate, or not at all, than one would expect from the phase space available, then there could be an inhibition from angular momentum conservation or some other conserved quantum number.

¹How to calculate dN/dE is covered in Chapter 2.

7.2 Weak interactions of leptons

Many of the key features of the weak interaction can be understood using leptonic decays and interactions without the additional complication of hadronic structure. Some examples of lepton properties that need explanation are the absence of the decay $\mu \rightarrow e\gamma$ (the upper limit on the branching ratio $\mu^- \rightarrow e^-\gamma$ is 1.2×10^{-11}), the very different lifetimes of the π^\pm (2.6×10^{-8} s) and π^0 (10^{-16} s), and the absence of purely leptonic decays such as $\mu^\pm \rightarrow e^\pm\nu$ or $\tau^\pm \rightarrow \mu^\pm\nu$. The much higher decay rate for the π^0 is straightforward: J^{PC} of the π^0 is 0^{-+} , so the electromagnetic decay $\pi^0 \rightarrow \gamma\gamma$ is allowed. So far, there is no evidence that any of the leptons are other than point-like particles.

7.2.1 Lepton number

The above brief summary of leptonic properties leads to the necessity for new conserved quantum numbers—the lepton numbers. For example, we assign $L_e = +1$ to the e^- and ν_e , $L_e = -1$ to the e^+ and $\bar{\nu}_e$, and $L_e = 0$ to all other particles. There are exactly analogous conservation laws for each of the other two generations involving L_μ and L_τ .² The details were summarized in Chapter 1, Table 1.1, which is repeated here for convenience as Table 7.1.

The lepton number conservation laws along with charge conservation and baryon number conservation are reflected in the fact that there is only a single vertex for each lepton generation that can be used in constructing Feynman diagrams for weak charged-current interaction. They are shown in Fig. 7.2.

7.2.2 Feynman rules

Calculations of high-energy interactions respecting Lorentz covariance and allowing for particle creation and annihilation are difficult even

State	Q	Mass	L_e	L_μ	L_τ	Lifetime
e^-	-1	0.511 MeV	+1	0	0	$> 4.6 \times 10^{26}$ years
ν_e	0	< 2 eV	+1	0	0	Stable
μ^-	-1	105.7 MeV	0	+1	0	$2.197034(21) \times 10^{-6}$ s
ν_μ	0	< 0.19 MeV	0	+1	0	Stable
τ^-	-1	1776.82 ± 0.16 MeV	0	0	+1	$(290.6 \pm 1.0) \times 10^{-15}$ s
ν_τ	0	< 18.2 MeV	0	0	+1	Stable

Table 7.1 Lepton properties.

²Although it was assumed at the time of its discovery that the τ lepton was a so-called sequential lepton with its own lepton number and an associated ν_τ , it was not until 1997 that the DONUT experiment confirmed directly that a ν_τ beam produced the charged τ .

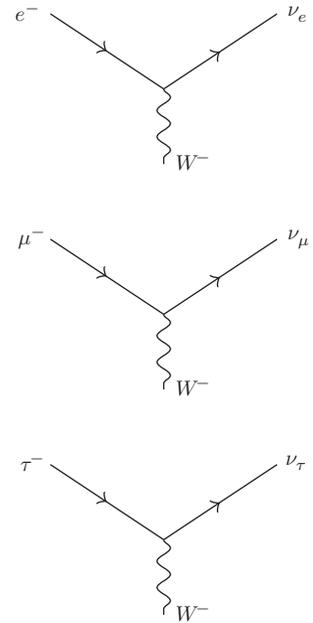


Fig. 7.2 Possible leptonic charged-current vertices.

for weak or electromagnetic interactions. Richard Feynman invented a very elegant graphical formalism that provides a considerable shortcut in calculations. The full mathematical treatment is given in the advanced textbooks in Further Reading, but a brief outline is given here since this formalism has become an essential part of the ‘language’ of particle physics—it will also provide a useful bridge to the more advanced texts. It ends up being formulated as a set of ‘Feynman rules’, in other words a recipe for performing the calculation. To show that Feynman rules actually work and give the right answer is even further beyond the scope of this book.

Feynman’s recipe is as follows. First, the process from incoming particles to outgoing particles is described in terms of ‘Feynman diagrams’ where lines represent ‘quanta’ or particles in momentum space. Lines representing the initial and final particles in an interaction must be ‘on-mass-shell’ particles, i.e. ones where $E^2 = p^2 + m^2$. Internal lines represent ‘virtual particles’, which do not have to satisfy the mass-shell constraint.³ Diagrams can have as many complicated lines as you like (but all connected), although with weak and electromagnetic interactions those with the fewest vertices (the lowest-order diagrams) are the most important. To do the calculation, all diagrams up to a given order must be considered and the amplitudes added together to get the matrix element. In this way, it is possible to have interference (constructive or destructive) between two diagrams.

The second step in Feynman’s recipe is to write down elements in a mathematical expression for each external line, vertex, and internal line (propagators) according to the rules, and the third step is to evaluate the expression. If there are internal loops, the evaluation involves integrating over the unconstrained momenta of those particles. The recipe ensures energy and momentum conservation and it is also built into the rules that the resulting expression is Lorentz-invariant.

Apart from being a precise mathematical tool for relativistic field theory calculations, Feynman diagrams give a very intuitive ‘billiard ball’ picture of what is a complicated interaction of waves and quanta. Other, now-commonplace ideas such as vertex factors (e.g. $\alpha = \frac{1}{137}$ for the electromagnetic interaction) and the propagators $1/q^2$ for a massless photon and $1/(M_W^2 - q^2)$ for the W for internal lines come from the Feynman rules.⁴

An example: tau decay

Using the Feynman rules, we will find the lowest-order matrix element for a purely leptonic decay mode of the tau lepton $\tau^- \rightarrow \nu_\tau e^- \bar{\nu}_e$ (Fig. 7.3). The Fermi expression for this process involves an integral over Schrödinger wavefunctions ψ :

$$M_{\text{fi}} = G_{\text{F}} \int \psi_{\nu_\tau}^* \psi_e^* \psi_{\bar{\nu}_e}^* \psi_\tau \, dV \quad (7.3)$$

³Although internal particles do not have to be on the mass shell, they do have to be consistent with the constraints of the time–energy uncertainty relation $\Delta E \Delta t \leq \hbar/2$.

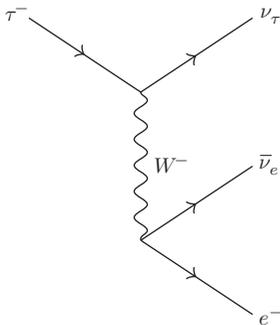


Fig. 7.3 Feynman diagram for a tau-lepton decay to an electron and two neutrinos.

⁴Note that the W width has been ignored here; it gives an imaginary part to the propagator, analogous to the expression for a Breit–Wigner amplitude—see Section 2.5.4.

The equivalent expression using Feynman rules and Dirac spinors is

$$M_{\text{fi}} = \frac{1}{M_W^2 - q^2} \left[\frac{g_w}{\sqrt{2}} \bar{u}(\nu_\tau) \gamma_\mu \frac{1}{2} (1 - \gamma_5) u(\tau) \right] \times \left[\frac{g_w}{\sqrt{2}} \bar{u}(e) \gamma^\mu \frac{1}{2} (1 - \gamma_5) u(\nu_e) \right] \quad (7.4)$$

where $1/(M_W^2 - q^2)$ is the propagator and the two expressions in square brackets are the vertex factors. This is an example of a ‘current–current’ interaction as described in Chapter 6. The vertex factors involve a particular way of combining the spinors—inserting a matrix $\gamma_\mu \frac{1}{2} (1 - \gamma_5)$ between the spinor and the adjoint spinor. This particular combination is called ‘V–A’—more details are given in Section 7.2.4. The factor g_w is called the weak coupling constant and is analogous to the electric charge $e = \sqrt{4\pi\alpha}$ (in natural units) in electromagnetism and q^2 in the propagator is the four-momentum squared of the W particle. The W does not have to be on the mass shell, since it is a virtual particle; however, since $q^2 \sim m_\tau^2$ and $m_\tau \ll M_W$, we can approximate $1/(M_W^2 - q^2) \rightarrow 1/M_W^2$, in which case we may use

$$\frac{G_F}{\sqrt{2}} = \frac{g_w^2}{8M_W^2} \quad (7.5)$$

to relate the weak coupling g_w to the Fermi constant G_F . The expression for the matrix element becomes

$$M_{\text{fi}} = \frac{4G_F}{\sqrt{2}} [\bar{u}(\nu_\tau) \gamma_\mu \frac{1}{2} (1 - \gamma_5) u(\tau)] [\bar{u}(e) \gamma^\mu \frac{1}{2} (1 - \gamma_5) u(\nu_e)] \quad (7.6)$$

which, apart from the γ matrices, now looks very much like the expression from Fermi theory in eqn 7.3. The factors of 4 and $\sqrt{2}$ are simply to keep the definition of G_F the same as originally defined by Fermi.

7.2.3 Universality

The model of weak interactions contains the postulate that

g_w is the same for *all* weak interactions . . .

. . . well, all weak charged-current interactions involving leptons. To bring quarks into the picture, we need Cabibbo theory and its extension by Kobayashi and Maskawa, and to include the neutral current, we need the Glashow, Weinberg, and Salam theory of electroweak unification.

There have been many experiments using a wide range of interactions and energies to test universality, measuring and comparing g_w . So far, the postulate that g_w is the same is holding up very well. We will discuss some of these experiments in Chapter 8.

Left-handed particles ...

You will recognize from Chapter 6 that the $\frac{1}{2}(1 - \gamma_5)$ in the operator is the left-hand chirality projection operator P_L (which, in the high-energy limit, is equivalent to a helicity projection). Denoting the left-handed part of the τ as $u_L(\tau)$, we have

$$u_L(\tau) = P_L u(\tau) = \frac{1}{2}(1 - \gamma_5)u(\tau) \quad (7.7)$$

This means that the $\frac{1}{2}(1 - \gamma_5)u(\tau)$ part of eqn 7.6 can be replaced by $u_L(\tau)$. We say that the weak charged-current interaction acts on only the left-handed part of the particle. We will formulate this more rigorously when we come to electroweak unification.

... and right-handed antiparticles

The $\frac{1}{2}(1 - \gamma_5)$ operator, when acting on the spinor of an antiparticle (using v rather than u to denote a spinor of an antiparticle), projects out the right-handed chirality of the antiparticle, i.e.

$$v_R = \frac{1}{2}(1 - \gamma_5)v \quad (7.8)$$

The weak charged current acts only on the left-handed part of particle wavefunctions and the right-handed part of antiparticle wavefunctions.

7.2.4 V–A

We now return to the question of why the Feynman rules were constructed so that the matrix between the spinors was $\gamma_\mu \frac{1}{2}(1 - \gamma_5)$. The first step is to construct a scalar quantity (i.e. a number that is Lorentz-invariant) out of a spinor u . The answer is that $\bar{u}u$ is such a quantity. It is the simplest of the ‘bilinear covariants’ for the Dirac equation and transforms as a Lorentz scalar. There are 16 such combinations⁵ and these can be arranged as shown in Table 7.2 to construct different types of Lorentz-covariant quantities. It can be shown that this is the only way

⁵See Section 7.3 of [80] for more details, starting from how a Dirac spinor transforms under a Lorentz transformation.

	Name		Number of parts	Parity P transformation
$\bar{u}u$	S	Scalar	1	+ under P
$\bar{u}\gamma_\mu u$	V	Vector	4	(+, −, −, −)
$\bar{u}\sigma_{\mu\nu}u$	T	Tensor	6	
$\bar{u}\gamma_\mu\gamma_5 u$	A	Axial vector	4	(−, +, +, +)
$\bar{u}\gamma_5 u$	P	Pseudoscalar	1	− under P

$$\sigma_{\mu\nu} = \frac{1}{2}i(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$$

Table 7.2 Lorentz covariant combinations of spinors.

to do it and this is the complete set of different combinations that can be made. It is also interesting for this discussion to see what happens under a parity transformation. The combination $\bar{u}u$ remains the same under a parity transformation. We can also construct other numbers by inserting various combinations of gamma matrices in the middle. For example, $\bar{u}\gamma_5 u$ is also a Lorentz-invariant scalar; however, in this case, the parity operation produces $-\bar{u}\gamma_5 u$, i.e. there is a change in sign. This is therefore called a ‘pseudoscalar’.

So we have a choice of S, V, T, A, or P (or some combination) for the interaction type for weak interactions. The second step is to turn to experiment to try to select the correct combination (see [57, pp. 398–401]). It involves (i) looking at Fermi ($\Delta J = 0$, where J is the total spin of the nucleus) and Gamow–Teller ($\Delta J = 0, 1$) allowed β decays and comparing various features of the decay electron spectra to see if the Fermi transitions are S or V or a combination of the two. The Gamow–Teller transitions tell us about the choice between A and T. After considerable experimental investigation, it was determined to be a combination of V and A, with no contributions from S or T (P cannot produce a large contribution). Further measurements gave the relative contributions of V and A (equal) and the sign (negative). So:

The weak interaction is V–A, i.e. $\gamma_\mu(1 - \gamma_5)$.

7.2.5 Parity violation

We have just decided that the weak interaction has a V–A form, i.e. that the spinor combinations $\bar{u}\gamma_\mu u$ and $\bar{u}\gamma_\mu\gamma_5 u$ appear with equal amounts. The first combination is a Lorentz 4-vector and the second is an axial vector. We use the V–A combination to describe the parity-violating weak interaction:

$$\bar{u}\gamma_\mu u - \bar{u}\gamma_\mu\gamma_5 u = \bar{u}\gamma_\mu(1 - \gamma_5)u \quad (7.9)$$

We now turn to the properties of these quantities under a parity operation. Both $\bar{u}\gamma_\mu u$ and $\bar{u}\gamma_\mu\gamma_5 u$ have definite parity transformation properties; however, they are opposite. If the weak interaction were a single type, i.e. either V or A, then there would not be parity violation—the final state would have definite parity. The V–A combination $\bar{u}\gamma_\mu(1 - \gamma_5)u$, however, has a mixture of parity change or non-change and this results in parity violation. Since the V and A parts come in equal amounts, parity is said to be maximally violated. We will review the experimental evidence for parity violation in the weak interaction in Chapter 8.

7.2.6 Currents and fields

We now give a reminder of the meaning of a current (see Chapter 6), since it will feature a lot in the electroweak theory. Consider the decay

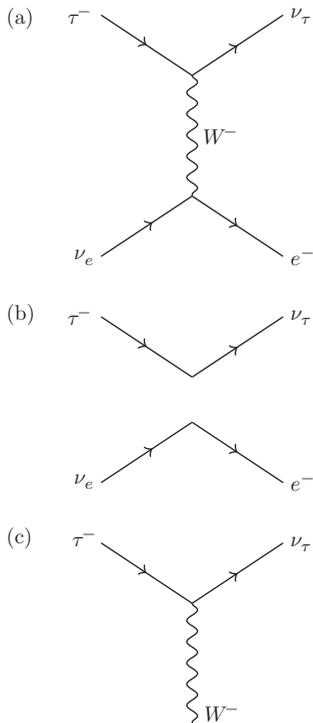


Fig. 7.4 (a) Feynman diagram for τ decay. (b) Same interaction showing the concept of two interacting currents. (c) Same, now showing the concept of a current interacting with a field.

of the τ , as shown in Fig. 7.4(a). The contents of the square brackets in eqn 7.6 are each referred to as currents:

$$J_\mu^+ = \bar{u}(\nu_\tau)\gamma_\mu \frac{1}{2}(1 - \gamma_5)u(\tau) \tag{7.10}$$

is a charge-raising current (because the outgoing ν_τ is one positive increment in charge bigger than the incoming τ^-). Similarly,

$$J^{-\mu} = \bar{u}(e)\gamma^\mu \frac{1}{2}(1 - \gamma^5)u(\nu_e) \tag{7.11}$$

is a charge-lowering current. These charge-raising (lowering) currents are the reason why this sort of interaction is called a charged current—the particle in the middle is a W^+ or a W^- and charge needs to be conserved at each vertex. The other type of weak interaction is the neutral current and involves the exchange of the Z^0 .

We can now write the matrix element in eqn 7.4 as

$$M_{\text{fi}} = \frac{g_w}{\sqrt{2}} J_\mu^+ \cdot \frac{1}{M_W^2 - q^2} \cdot \frac{g_w}{\sqrt{2}} J^{-\mu} \tag{7.12}$$

This is a so called current–current formalism of an interaction; i.e. the interaction occurs between the two currents, as shown in Fig. 7.4(b).

Another variation on this theme is to treat the two halves of the Feynman diagram separately. We can ‘plug-and-play’, i.e. change the particles at one vertex (e.g. from (e, ν_e) to (d, u)) without changing the calculation at the other vertex. We do this by introducing the concept of a field W^μ ,

$$W^\mu = \frac{1}{M_W^2 - q^2} \cdot \frac{g_w}{\sqrt{2}} J^{-\mu} \tag{7.13}$$

and the matrix element is

$$M_{\text{fi}} = \frac{g_w}{\sqrt{2}} J_\mu^+ \cdot W^\mu \tag{7.14}$$

This is shown in Fig. 7.4(c).

Another example of a current \times field interaction is the electromagnetic interaction. The equivalent current for an electromagnetic interaction (e.g. for an electron, with negative charge) is

$$J_\mu^{\text{EM}} = -\bar{u}(e)\gamma_\mu u(e) \tag{7.15}$$

and the electromagnetic field is A^μ , giving⁶ $M_{\text{fi}} = J_\mu^{\text{EM}} \cdot A^\mu$.

This discussion in terms of currents and fields is reminiscent of electric currents and magnetic fields—a quantity called a magnetic field is constructed at all points in space to describe the action of all the bits of current that are flowing in all circuits nearby. This can then be used to determine the force on a particular piece of current placed somewhere⁷.

⁶Note the absence of the projection operator because the photon couples with equal strength to the left-handed and right-handed parts.

⁷As we showed in Chapter 6, the continuity equation for the Dirac equation gives the expression for the Dirac spinor equivalent of a probability current as $\bar{u}\gamma_\mu u$.

7.3 Weak interactions including quarks

Having understood the $V-A$ structure of the weak interaction in the lepton sector, we are now ready to extend the idea of ‘universality’ to the strong sector and quarks. We shall see that to maintain universality it will be necessary to accept that the quark eigenstates involved in the weak interaction are linear combinations of those involved in strong interactions. This requires the introduction of Cabibbo theory and its extension to the CKM matrix. The result is a simple picture that explains a large number of experimental measurements. The background to the problem almost predates accelerator particle physics, going back to a time when the highest-energy data came from cosmic-ray physics. It was noted that some of the particles behaved in a strange way—after correction for phase space, they seemed to decay about a factor of 20 more slowly than those that were ‘not strange’.⁸

⁸The ‘slow ones’ turn out to contain a strange quark—hence the quark name.

7.3.1 Cabibbo theory

Figure 7.5(a) shows the quarks arranged in families with the possible charged-current transitions marked (thicker lines denote the transitions with the highest probabilities). In contrast, Fig. 7.5(b) shows the lepton families. Extra transitions are required to convert quarks of different generations, which do not exist in the interactions of the leptons. If this were not the case, the K and Λ particles would be completely stable (as would some particles with b quarks). Cabibbo developed his theory when only the u , d , and s quarks were known. The theory states:

The charged-current interactions of quarks proceed with the same coupling constant g_w as for leptonic interactions provided we associate a factor $\cos\theta_C$ to $u \leftrightarrow d$ transitions and a factor $\sin\theta_C$ to $u \leftrightarrow s$ transitions.

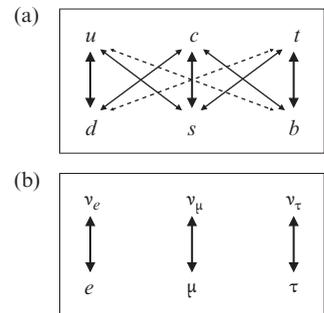


Fig. 7.5 Allowed charged-current interactions of quarks (a) and leptons (b).

With this scheme, a large number of weak interactions and decays are correctly predicted; the value of θ_C required is 13.1° .

However, a remaining problem was the decay $K_L^0 \rightarrow \mu^+ \mu^-$, a second-order weak interaction requiring two W particles in the lowest-order Feynman diagram. The predicted rate was many orders of magnitude faster than experimentally measured. A solution to this paradox was suggested by Glashow, Iliopoulos, and Maiani (GIM) [78].

7.3.2 GIM mechanism, flavour-changing neutral currents

The GIM mechanism introduced a fourth quark c into the theory,⁹ with couplings $\cos\theta_C$ for $c \leftrightarrow s$ transitions and $-\sin\theta_C$ for $c \leftrightarrow d$. This produces a second Feynman diagram in the process $K_L^0 \rightarrow \mu^+ \mu^-$, which interferes destructively with the one involving only u , d , and s quarks and suppresses the decay to a level compatible with experiment (see

⁹This was in 1970, four years before the J/ψ was discovered.

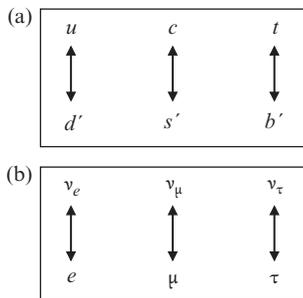


Fig. 7.6 Allowed charged-current interactions with Cabibbo-rotated quarks (a) and leptons (b).

[80, p. 327] or [84, p. 282]), provided the mass difference $m_c - m_u$ is not too large.

With four quarks, the GIM mechanism gives a nice conceptual picture of what is going on. If we define a linear superposition of quarks (of flavour eigenstates d and s) to form d' and s' ,

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos \theta_C & \sin \theta_c \\ -\sin \theta_C & \cos \theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \quad (7.16)$$

then it is possible to view weak charged-current interactions of quarks as occurring within generations with the same coupling g_w as for leptons, provided we consider the rotated states d' and s' rather than d and s . This is shown in Fig. 7.6(a) (ignore the t and b quarks in the diagrams for the moment).

The original motivation for the GIM mechanism was to suppress weak interactions that change flavour but not the quark charge (flavour-changing neutral currents), such as $s \rightarrow d$, which are only observed at very low rates. The GIM mechanism will show that these interactions are forbidden at tree level and it also ensures that the rates are suppressed at higher order in perturbation theory.

The idea is as follows. The Cabibbo rotation in eqn 7.16 shows that we have to consider weak charged-current interaction as being between quarks $u \leftrightarrow d'$ and $c \leftrightarrow s'$ and that these form two families that do not mix at all, just like the leptons (ignoring neutrino oscillations). Therefore, for the weak neutral-current interaction, we can hypothesize that the same thing happens, i.e. that the quark families do not mix at all in the rotated basis; the only interactions allowed are $u \leftrightarrow u$, $d' \leftrightarrow d'$, $s' \leftrightarrow s'$ and $c \leftrightarrow c$.

The question is: restricting ourselves to the above hypothesis, are flavour-changing interactions of the neutral current, i.e. $d \leftrightarrow s$ allowed? If so, then the overlap between d and s should not be zero. Take the rotation defined in eqn 7.16, invert it, and use it to give

$$\begin{aligned} (ds) &= (d' \cos \theta_C - s' \sin \theta_C)(d' \sin \theta_C + s' \cos \theta_C) \\ &= (d'd' - s's') \cos \theta_C \sin \theta_C + d's' \cos^2 \theta_C - s'd' \sin^2 \theta_C \\ &= 0 \end{aligned} \quad (7.17)$$

where the last line comes from the hypothesis that in the new basis, there are no interactions between quark families, i.e. $d'd' = s's' = 1$ and $d's' = s'd' = 0$. So lowest-order flavour-changing neutral currents are not possible under the GIM mechanism. This then agrees with experiment, since flavour-changing neutral currents are observed to be highly suppressed.¹⁰

¹⁰Flavour-changing neutral currents are allowed at higher order and lead to the very important flavour oscillations such as $K^0 \rightarrow \bar{K}^0$, which we will examine in Chapter 10.

7.3.3 CKM matrix

As hinted by including t and b quarks in Figs. 7.5 and 7.6, the scheme can be extended to three quark generations with a 3×3 matrix, which was proposed by Kobayashi and Maskawa (the resulting mixing matrix \mathbf{V} is known as the CKM matrix, where C = Cabibbo):

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \mathbf{V} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (7.18)$$

The matrix has values that come from experiment. The magnitudes (we will consider the phases later) are the magnitudes (...) are as follows:¹¹

¹¹For the most recent values consult the PDG Tables [115].

$$\mathbf{V} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} \quad (7.19)$$

We will review some of the techniques used to measure the elements of the CKM matrix in Chapter 8. Note that the bottom row and right column are rather close to being all zero or one, i.e. the third generation does not mix much with the other two. This is the reason why the decays of particles involving b quarks are very slow.

Kobayashi and Maskawa were originally searching for an excuse to allow a non-trivial complex number into the 2×2 mixing matrix—they realized¹² that this is a way of inserting (at that time recently discovered) CP violation into the theory. Because it is possible to add a phase to each quark without altering the theory (measurable quantities are proportional to $|M|^2$), this could not be done with a 2×2 matrix. If the number of quark generations is extended to 3, the matrix can have one non-trivial CP violation-generating phase. Much of what is known as ‘heavy-flavour physics’ revolves around pinning down the values of the CKM matrix elements, since they are key parameters in calculating the decay rates of heavy mesons and baryons. In addition to measurements depending directly on the matrix elements, the CKM matrix must also respect the constraint of unitarity. Although the number of parameters to be determined is not huge, to take proper account of the mathematical constraints and to include systematic and statistical errors correctly is non-trivial and beyond the scope of this text.

¹²For which they won the 2008 Nobel Prize.

It is conventional to treat the charge $-\frac{1}{3}$ quarks (d, s, b) as the ones that mix to (d', s', b'), while the others (u, c, t) do not change. This could have been done the other way round (or even considering a combination of rotations of both the charge $-\frac{1}{3}$ and charge $+\frac{2}{3}$ quarks), but it can be shown to simplify to the same combinations as used here.

Assuming that the CKM matrix is unitary, it can be parameterized in terms of three independent mixing angles θ_{12} , θ_{23} , θ_{13} and the one complex phase δ as discussed above. A popular way to parameterize the matrix, following the Particle Data Group (PDG), is as follows:

$$\mathbf{V} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \quad (7.20)$$

Here $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$, where $i, j = 1, 2, 3$ are generation labels. In the limit $\theta_{23} = \theta_{13} = 0$, the third generation decouples from the first two and $\theta_{12} = \theta_C$ (the Cabibbo angle). The phase δ allows for CP violation in the standard model. CP violation will be covered further in Chapter 10.

7.3.4 Decays of hadrons containing heavy quarks

We are now in a position to try to make predictions of how weak decays of mesons and baryons containing heavy quarks might proceed. A useful concept is that of the ‘spectator quark’. Normally during the existence of a meson, gluon exchange occurs continuously between the $q\bar{q}'$ pair, so the strong force is important in understanding what is happening. The idea is that when the hadron decays, one of its constituent quarks changes flavour, emitting a virtual W particle—while this is happening, the other quark(s) are ‘spectators’, i.e. they play no role in the weak interaction itself. Although this is clearly not a rigorous result, it does provide a useful approximate model. An example is shown in Fig. 7.7, where the \bar{d} is a spectator quark during the decay $\bar{K}^0 \rightarrow \pi^+\pi^-$.

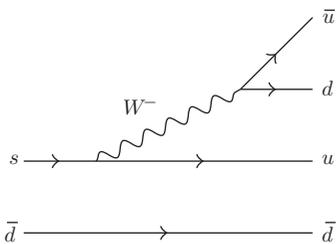


Fig. 7.7 Spectator quark diagram for $\bar{K}^0 \rightarrow \pi^+\pi^-$ decay.

From inspecting the CKM matrix in eqn 7.19, we can see that since V_{cs} is large, a c quark within a meson is going to decay preferentially to an s quark. For example, the decay $D^+ \rightarrow K + \text{anything}$ should have a high branching ratio, and indeed such decays dominate: $K^+ + \text{anything}$ 28%; $K^0/\bar{K}^0 + \text{anything}$ 61%; $K^- + \text{anything}$ 5.5%. Another indication that the spectator model is correct would be if the lifetime of a charged D meson were the same as the lifetime of a neutral D . This is not quite the case: $\tau(D^+) = 1.040$ ps and $\tau(D^0) = 0.410$ ps. However, the spectator-quark model ignores some other effects: the D^0 has more annihilation diagrams than the D^+ and non-perturbative effects from the strong interaction are still significant.

¹³Which will subsequently decay to mesons with an s quark.

The B mesons are expected to decay mostly to particles involving charm,¹³ because V_{cb} is about a factor of 10 bigger than V_{ub} , and this is indeed the case. Since neither V_{cb} or V_{ub} is very big, the b quark decays relatively slowly. The measured lifetimes are $\tau(B^+) = 1.67$ ps and $\tau(B^0) = 1.54$ ps, so the spectator-quark model prediction is much better. The larger mass of B mesons means that $\alpha_s(m_b)$ is smaller and perturbative effects are less important (this is a consequence of ‘running’ coupling constants—see Chapter 9).

Looking again at the CKM matrix, we see that V_{tb} is nearly 1. This means that the t quark decays nearly always to the b quark and not directly to s or d quarks. Since it is very heavy, the phase space is very large and it will decay rapidly—indeed too rapidly for any meson to be formed.

A brief reminder about the naming of heavy mesons: the letter B or D with no subscript means the meson contains a heavy quark and either a u or a d in order to make up the charge of the meson. More exotic mesons are denoted with a subscript that indicates the less massive quark, e.g. B_s , D_s , or even B_c .

7.4 Introduction to electroweak unification

In Sections 7.2 and 7.3, we have characterized the charged-current weak interaction of both quarks and leptons as follows:

- The interaction proceeds with a coupling constant g_w for all weak charged-current processes. $G_F/\sqrt{2} = g_w^2/8M_W^2$.
- The quark flavour eigenstates are rotated through the Cabibbo angle $\theta_C \simeq 13.1^\circ$ (or for six quarks by the 3×3 CKM matrix).
- The ‘vertex factor’ is found to be (e.g. for $u \rightarrow d$) $g_w V_{ud} J_\mu^-$, where the weak charged current $J_\mu^- = \bar{u}(u)\gamma_\mu \frac{1}{2}(1 - \gamma_5)u(d)$ and V_{ud} is the appropriate element of the CKM matrix.

By comparing the properties of the electromagnetic (EM) and weak forces, it becomes apparent that these have some very similar properties. Perhaps it might be possible to unify them, i.e. to provide a theory that covers both forces as two aspects of a more complete theory. This was achieved by Glashow, Salam, and Weinberg (GSW) and we will go through their arguments here. The theory was constructed before neutral currents had been discovered, and indeed it predicted the properties of neutral currents, which were subsequently experimentally verified. We will start with just the EM and weak charged-current interactions and see how the neutral current emerges from the theory.

First, in Table 7.3, we list the properties of the forces more carefully. Although the forces have similarities, they are clearly different. The procedure used by GSW to unify the forces is divided into four steps, since it is complicated to explain. We will use the $\tau^- \rightarrow \nu_\tau e^- \bar{\nu}$ decay from Section 7.2.2 as an example.

7.4.1 Electroweak unification procedure

The process of electroweak unification starts with the components of the weak interaction as they were known around 1964 (i.e. only the postulated W^\pm , and no neutral weak interaction), and the success of the Feynman diagram view of perturbation theory for QED calculations. In exactly the same way that local gauge invariance in the Dirac equation leads to an additional spin-1 field (the photon) and the correct photon–electron interaction (see Section 6.5), we use the same mechanism to insert the weak charged-current interaction of the fermions by a spin-1 boson (the W^\pm).¹⁴ We assume the V–A structure from Section 7.2.4 and that the spin of the W is 1, as guided by experiment, so the current is as in Section 7.2.6:

$$J_\mu^+ = \bar{u}(\nu)\gamma_\mu \frac{1}{2}(1 - \gamma_5)u(\tau) \quad (7.21)$$

¹⁴We ignore for now the fact that the local gauge invariance mechanism only works with massless particles like photons—we will return to this later.

EM interaction	Weak CC interaction
Maxwell's equations	e.g. $\nu_e n \rightarrow p e^-$, $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$, $K^0 \rightarrow e^+ \pi^- \nu_e$
Long range	Short range
γ = massless spin-1 boson	Massive W^\pm spin-1 bosons
$F = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^2}$	Propagator $\frac{1}{M_W^2 - q^2}$
Acts on particles depending on their charge	Acts on left-handed particles and right-handed antiparticles only
Conserves parity	Not parity-conserving
Comes from the Dirac equation when insisting on local gauge invariance	
$J_\mu^{\text{EM}} = \bar{\psi} \gamma_\mu Q \psi$ where Q is the electric charge in units in which the electron has $Q = -1$	$J_\mu = \bar{\psi} \gamma_\mu \frac{1}{2}(1 - \gamma_5) \psi$
Coupling constant $e = \sqrt{4\pi\alpha}$	Coupling constant g_w for all weak interactions (universality); use CKM matrix elements with quarks

Table 7.3 Comparison of properties of the electromagnetic and weak charged-current interactions.

Step 1: Left-handed particles

Since $P_L = \frac{1}{2}(1 - \gamma_5)$ is a projection operator, we can operate on a spinor with it twice in succession without changing the effect, so $\frac{1}{2}(1 - \gamma_5)u(\tau) = \frac{1}{2}(1 - \gamma_5)\frac{1}{2}(1 - \gamma_5)u(\tau)$, and

Particles	Antiparticles
$u_L = \frac{1}{2}(1 - \gamma_5)u$	$v_L = \frac{1}{2}(1 + \gamma_5)v$
$u_R = \frac{1}{2}(1 + \gamma_5)u$	$v_R = \frac{1}{2}(1 - \gamma_5)v$
$\bar{u}_L = \bar{u}\frac{1}{2}(1 + \gamma_5)$	$\bar{v}_L = \bar{v}\frac{1}{2}(1 - \gamma_5)$
$\bar{u}_R = \bar{u}\frac{1}{2}(1 - \gamma_5)$	$\bar{v}_R = \bar{v}\frac{1}{2}(1 + \gamma_5)$

$$J_\mu^+ = \bar{u}(\nu)\gamma_\mu \frac{1}{2}(1 - \gamma_5)\frac{1}{2}(1 - \gamma_5)u(\tau) \tag{7.22}$$

$$= \bar{u}(\nu)\frac{1}{2}(1 + \gamma_5)\gamma_\mu \frac{1}{2}(1 - \gamma_5)u(\tau) \tag{7.23}$$

$$= \left[\bar{u}(\nu)\frac{1}{2}(1 + \gamma_5) \right] \gamma_\mu \left[\frac{1}{2}(1 - \gamma_5)u(\tau) \right] \tag{7.24}$$

$$= \bar{u}(\nu_L)\gamma_\mu u(\tau_L) \tag{7.25}$$

R, L correspond to helicity +1, -1 if $m=0$ and approximately if $m \sim 0$.

Table 7.4 Spinors of particles and antiparticles.

where in line 7.23, we have used the identity $\gamma_\mu \gamma_5 + \gamma_5 \gamma_\mu = 0$.

Line 7.24 is the main part of step 1 towards unification—we associate the parts with the $(1 \pm \gamma_5)$ with the spinors rather than the operator. By doing so, we are left with an operator γ_μ that looks like the EM operator. This looks more obvious in line 7.25, where we write the left-handed parts of the spinors directly. Table 7.4 shows the correspondence between

left- and right-handed spinors and adjoint spinors for both particles and antiparticles.

W^\pm vertex factors are the same as EM if we act *only* on the left-handed part of the spinor.

Interlude: weak isospin

We now add some structure to formalize step 1 by adding a quantity I_3 that plays an analogous role for the weak interaction as the charge Q does for the EM interaction. We categorize the particles—indeed the separate left- and right-handed parts¹⁵—according to how they interact weakly. We define a quantity called *weak isospin* in a mathematically analogous way to isospin (and ordinary spin), and arrange the left-handed parts of the particles in $I = \frac{1}{2}$ doublets such as $(\nu_{\mu L}, \mu_L)$ or (u_L, d'_L) containing the particles that can change into each other at a weak CC vertex. The member of each multiplet with the more positive charge is assigned $I_3 = +\frac{1}{2}$ and the member with the more negative charge is assigned $I_3 = -\frac{1}{2}$.

¹⁵The left- and right-handed parts are not particles in their own right. An electron spinor $u(e)$ describes the real electron; we split it up as $u(e) = u(e_L) + u(e_R)$ simply to make the similarity between the EM and weak interactions apparent in the formalism.

I	I_3						
$\frac{1}{2}$	$+\frac{1}{2}$	ν_{eL}	$\nu_{\mu L}$	$\nu_{\tau L}$	u_L	c_L	t_L
$\frac{1}{2}$	$-\frac{1}{2}$	e_L	μ_L	τ_L	d'_L	s'_L	b'_L

The right-handed parts of all the particles are assigned to weak isospin singlets with $I = 0$ and $I_3 = 0$.

I	I_3									
0	0	e_R	μ_R	τ_R	u_R	c_R	t_R	d'_R	s'_R	b'_R

The right-handed neutrinos have been left out of the table—their nature is still being investigated. They also have $I_3 = 0$. It is possible that neutrinos are Majorana particles (see Chapter 11), i.e. they are their own antiparticles (in which case, flipping the helicity of a neutrino ν_L produces the antineutrino $\bar{\nu}_R$; neutrinoless double β decay becomes possible), or they could be Dirac particles in which neutrino and antineutrino are distinct, and therefore the ν_R and $\bar{\nu}_L$ do not interact with any known force. If neutrinos had been exactly massless, these two situations would be experimentally indistinguishable.

The I_3 label, which is $+\frac{1}{2}$, $-\frac{1}{2}$, or 0 is used for the weak charged-current interaction in a similar way to the charge in an electromagnetic interaction—i.e. when it is 0, there is no interaction.

Step 2: W^0

We next appeal to symmetry, and postulate a neutral partner W^0 to the W^\pm . This interacts producing no change in charge within each doublet.

- It has the same coupling constant g_w .
- It interacts only with the left-handed states.

This is *not* the Z^0 , which also interacts with right-handed states of quarks and leptons.

Step 3: B^0

Instead of the electromagnetic interaction, we introduce another field, the B^0 , which will ensure the correct electromagnetic interaction in the following step. The B^0 interacts with a strength proportional to a quantity called weak hypercharge Y , where Y is defined by

$$Q = I_3 + \frac{1}{2}Y \quad (7.26)$$

The B^0 interacts with a current J_μ^Y :

$$J_\mu^Y = 2J_\mu^{\text{EM}} - 2J_\mu^0 \quad (7.27)$$

We give the B^0 interaction its own coupling constant $g'/2$ (the factor of 2 is just a convention). Examples of B^0 currents involving left-handed up quarks and right-handed electrons are

$$J_\mu^Y = \bar{u}(u_L)\gamma_\mu Y(u_L)u(u_L), \quad \bar{u}(e_R)\gamma_\mu Y(e_R)u(e_R) \quad (7.28)$$

where $Y(u_L) = +\frac{1}{3}$ and $Y(e_R) = -2$ using eqn 7.26. Table 7.6 at the end of this section gives Y for all the particles.

Step 4: The Weinberg angle

We will use the notation introduced in Section 7.2.6 with currents and fields. Table 7.5 gives a summary of the symbols used, the same as those above with the addition of μ indices on the currents and fields. Since all the fields are vector or axial vector fields, they include an index $\mu = 0, 1, 2, 3$. When a current and a field are combined to form a matrix element, there is an implied summation over μ .

The electromagnetic interaction A^μ is a linear combination of $W^{\mu,0}$ and B^μ , and the orthogonal combination produces a new interaction, which is the weak neutral current Z^0 . A convenient way to form the linear combination is with a rotation angle. We introduce a new rotation angle θ_W , the ‘weak mixing angle’ or ‘Weinberg angle’.

$$\begin{aligned} A^\mu &= +B^\mu \cos \theta_W + W^{\mu,0} \sin \theta_W \\ Z^\mu &= -B^\mu \sin \theta_W + W^{\mu,0} \cos \theta_W \end{aligned} \quad (7.29)$$

	Current	Field	Coupling constant	'Charge'
Charge-raising weak	J_μ^+	$W^{\mu,+}$	g_w	I_3
Charge-lowering weak	J_μ^-	$W^{\mu,-}$	g_w	I_3
Symmetric W^0 field	J_μ^0	$W^{\mu,0}$	g_w	I_3
B^0 field (step 3)	J_μ^Y	B^μ	$g'/2$	Y
Electromagnetic	J_μ^{EM}	A^μ	e	Q
Neutral current	J_μ^{NC}	Z^μ		

Table 7.5 Summary of fields discussed in this section and the symbols used.

We now write down the total GSW electroweak interaction in the form of a (current) $_\mu$ (field) $^\mu$ Lorentz scalar:¹⁶

$$g_w \left(J_\mu^+ \frac{W^{\mu,+}}{\sqrt{2}} + J_\mu^- \frac{W^{\mu,-}}{\sqrt{2}} + J_\mu^0 W^{\mu,0} \right) + \frac{g'}{2} J_\mu^Y B^\mu \quad (7.30)$$

¹⁶A more detailed treatment constructs the W^+ and W^- in a way that makes the symmetry with W^0 more apparent—see Further Reading.

The factors of $\sqrt{2}$ come from group theory. Now, we invert eqn 7.29,

$$B^\mu = +A^\mu \cos \theta_W - Z^{\mu,0} \sin \theta_W \quad (7.31)$$

$$W^\mu = +A^\mu \sin \theta_W + Z^{\mu,0} \cos \theta_W \quad (7.32)$$

and write the neutral part of eqn 7.30 in terms of A^μ and Z^μ by substituting in from eqns 7.31 and 7.32:

$$g_w J_\mu^0 W^{\mu,0} + \frac{g'}{2} J_\mu^Y B^\mu = \left(g_w \sin \theta_W J_\mu^0 + g' \cos \theta_W \frac{J_\mu^Y}{2} \right) A^\mu + \left(g_w \cos \theta_W J_\mu^0 - g' \sin \theta_W \frac{J_\mu^Y}{2} \right) Z^\mu \quad (7.33)$$

Recovering the EM interaction

Consider next the two parts of the right-hand side of eqn 7.33 separately. The first part must be set equal to $e J_\mu^{\text{EM}} A^\mu$, otherwise the GSW theory will not reproduce the EM physics described by QED. From eqn 7.27, $J_\mu^{\text{EM}} = J_\mu^0 + \frac{1}{2} J_\mu^Y$, and so

$$\left(g_w \sin \theta_W J_\mu^0 + g' \cos \theta_W \frac{J_\mu^Y}{2} \right) A^\mu = e \left(J_\mu^0 + \frac{1}{2} J_\mu^Y \right) A^\mu \quad (7.34)$$

For this to be satisfied, we must set

$$e = g_w \sin \theta_W = g' \cos \theta_W \quad (7.35)$$

These two equalities are called the ‘unification condition’.

The neutral-current interaction

The second part of eqn 7.33 is the weak neutral current $g_z J_\mu^Z$. This is completely specified (i.e. there are no free parameters) by the charged weak current and electromagnetism:

$$g_z J_\mu^Z = g_w \cos \theta_W J_\mu^0 - g' \sin \theta_W \frac{J_\mu^Y}{2} \quad (7.36)$$

We now manipulate this expression to remove the parts involving B^0 . Using eqn 7.27 and the unification condition 7.35, we have

$$\begin{aligned} g_z J_\mu^Z &= \frac{g_w}{\cos \theta_W} [\cos^2 \theta_W J_\mu^0 - \sin^2 \theta_W (J_\mu^{\text{EM}} - J_\mu^0)] \\ &= \frac{g_w}{\cos \theta_W} (J_\mu^0 - \sin^2 \theta_W J_\mu^{\text{EM}}) \end{aligned} \quad (7.37)$$

We can now put explicit forms of $J_\mu^0 = I_3 \bar{u} \gamma_\mu \frac{1}{2} (1 - \gamma_5) u$ and $J_\mu^{\text{EM}} = Q \bar{u} \gamma_\mu u$ into this expression to give

$$\begin{aligned} g_z J_\mu^Z &= \frac{g_w}{\cos \theta_W} \bar{u} \gamma_\mu \left[\frac{1}{2} (1 - \gamma_5) I_3 - \sin^2 \theta_W Q \right] u \\ &= \frac{g_w}{\cos \theta_W} \bar{u} \gamma_\mu \left[\frac{1}{2} (1 - \gamma_5) I_3 - \sin^2 \theta_W Q \left(\frac{1}{2} (1 - \gamma_5) + \frac{1}{2} (1 + \gamma_5) \right) \right] u \\ &= \frac{g_w}{\cos \theta_W} \bar{u} \gamma_\mu \left[g_L \frac{1}{2} (1 - \gamma_5) + g_R \frac{1}{2} (1 + \gamma_5) \right] u \end{aligned} \quad (7.38)$$

where $g_L = I_3 - Q \sin^2 \theta_W$ and $g_R = -Q \sin^2 \theta_W$ are the couplings of the left- and right-handed particles, respectively. Note that, apart from neutrinos for which $Q = 0$ and so $g_R = 0$, the neutral current interacts with both the left- and right-handed states of the particle but with different strengths. In contrast, the charged-current interaction involves only the left-handed states of all particles.

It is also possible to rearrange eqn 7.38 to define coupling constants c_V and c_A in terms of g_L and g_R :

$$g_z J_\mu^Z = \frac{g_w}{\cos \theta_W} \bar{u} \gamma_\mu \left[\frac{1}{2} (g_L + g_R) - \frac{1}{2} (g_L - g_R) \gamma_5 \right] u \quad (7.39)$$

$$= \frac{g_w}{\cos \theta_W} \bar{u} \gamma_\mu \frac{1}{2} (c_V - c_A \gamma_5) u \quad (7.40)$$

Fermion (f)	I	I_3	Q	Y	c_V^f	c_A^f
$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}$	$\frac{1}{2}$	$+\frac{1}{2}$ $-\frac{1}{2}$	0 -1	-1 -1	$+\frac{1}{2}$ $-\frac{1}{2} + 2\sin^2\theta_W$	$+\frac{1}{2}$ $-\frac{1}{2}$
$\nu_{eR}, \nu_{\mu R}, \nu_{\tau R}$	0	0	0	0		
e_R, μ_R, τ_R	0	0	-1	-2		
$\begin{pmatrix} u_L \\ d'_L \end{pmatrix}, \begin{pmatrix} c_L \\ s'_L \end{pmatrix}, \begin{pmatrix} t_L \\ b'_L \end{pmatrix}$	$\frac{1}{2}$	$+\frac{1}{2}$ $-\frac{1}{2}$	$+\frac{2}{3}$ $-\frac{1}{3}$	$+\frac{1}{3}$ $+\frac{1}{3}$	$+\frac{1}{2} - \frac{4}{3}\sin^2\theta_W$ $-\frac{1}{2} + \frac{2}{3}\sin^2\theta_W$	$+\frac{1}{2}$ $-\frac{1}{2}$
u_R, c_R, t_R	0	0	$+\frac{2}{3}$	$+\frac{4}{3}$		
d_R, s_R, b_R	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$		

Table 7.6 The electroweak properties associated with each group of fermions.

where $c_V = I_3 - 2Q\sin^2\theta_W$ and $c_A = I_3$. Different applications prefer to use either g_L, g_R or c_V, c_A .¹⁷ The values of c_V^f and c_A^f for each fermion f along with other properties of the fermions are shown in Table 7.6.

7.4.2 Weak neutral currents

The existence of weak neutral currents (see Section 8.3) was the first critical prediction of the unified electroweak theory. The amplitude for $\bar{\nu}_\mu e \rightarrow \bar{\nu}_\mu e$ is

$$M = \frac{g_w^2}{8M_Z^2 \cos^2\theta_W} [\bar{u}(\bar{\nu}_\mu)\gamma_\mu(c_V^{(\nu)} - c_A^{(\nu)}\gamma_5)u(\bar{\nu}_\mu)] \quad (7.41)$$

$$\times [\bar{u}(e)\gamma^\mu(c_V^{(e)} - c_A^{(e)}\gamma_5)u(e)]$$

We can look up the couplings from Table 7.6, which are $c_V^{(\nu)} = c_A^{(\nu)} = \frac{1}{2}$, $c_V^{(e)} = -\frac{1}{2} + 2\sin^2\theta_W$, and $c_A^{(e)} = -\frac{1}{2}$, and so

$$M = \frac{g_w^2}{8M_Z^2 \cos^2\theta_W} \left[\bar{u}(\bar{\nu}_\mu)\gamma_\mu \frac{1}{2}(1 - \gamma_5)u(\bar{\nu}_\mu) \right] \quad (7.42)$$

$$\times \left[\bar{u}(e)\gamma^\mu \left(2\sin^2\theta_W - \frac{1}{2}(1 - \gamma_5) \right) u(e) \right]$$

¹⁷For simplicity, we can use the value of I_3 for the left-handed particle in these equations all the time; if we have a right-handed component to the particle, we can write it as $\frac{1}{2}(1 + \gamma_5)u$, and when combined with the $(1 - \gamma_5)$ in eqn 7.38, it causes the term with I_3 in it to be zero.

Note that the current involving the neutrino takes the same $\frac{1}{2}(1 - \gamma_5)$ form as the charged current. What is special about the neutrino that causes this? It is the fact that it is electrically neutral, so all of its interaction comes from the W^0 , with none from the B^0 .

The calculation of the cross section involves averaging over initial spins and summing over final spins. It can be done with several time-saving tricks (see Further Reading). The result for $d\sigma(\bar{\nu}_\mu e \rightarrow \bar{\nu}_\mu e)/dE_e$ is

$$\frac{d\sigma}{dE_e} = \frac{G_F^2 m_e}{2\pi} \left\{ \left(c_V^{(e)} - c_A^{(e)} \right)^2 + \left(c_V^{(e)} + c_A^{(e)} \right)^2 \left(1 - \frac{E_e}{E_\nu} \right)^2 - \frac{m_e E_e}{E_\nu^2} \left[\left(c_V^{(e)} \right)^2 - \left(c_A^{(e)} \right)^2 \right] \right\} \quad (7.43)$$

This can be integrated to give the cross section

$$\sigma = \int \frac{d\sigma}{dE_e} dE_e \sim E_\nu \times 10^{-45} \text{ m}^2$$

(where E_ν is in GeV)—a very small cross-section! The neutral current is also observable in neutrino–nucleon collisions. The cross sections are still small, but somewhat larger than for scattering off electrons. The observation and precision measurements of neutral currents will be discussed in Chapter 8.

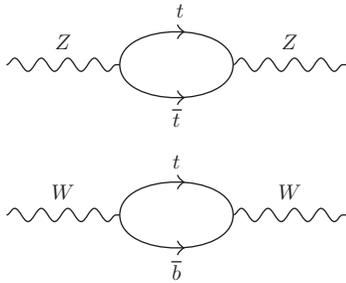


Fig. 7.8 Radiative corrections to the W and Z masses from top-quark loops.

7.4.3 Masses of W and Z bosons

We can now use the electroweak unification theory to predict the masses of the W and Z bosons in terms of the weak mixing angle $\sin \theta_W$ and the Fermi coupling constant G_F . We start with the relation between G_F and M_W , eqn 7.5, and, substituting for g_W from eqn 7.35, we find

$$\frac{G_F}{\sqrt{2}} = \frac{e^2}{8M_W^2 \sin^2 \theta_W} \quad (7.44)$$

$$M_W = \left(\frac{\sqrt{2}e^2}{8G_F} \right)^{1/2} \frac{1}{\sin \theta_W}$$

Using the unification condition 7.35 again, we can simply relate the masses of the W and Z bosons in terms of the weak mixing angle:

$$\frac{M_W}{M_Z} = \cos \theta_W \quad (7.45)$$

Therefore, if we have measurements of $\sin^2 \theta_W$ and G_F from low-energy experiments, we can predict the masses of the W and the Z bosons. Hence the discovery of the W and the Z bosons (see Chapter 8) at the expected masses was a triumph for the electroweak theory.

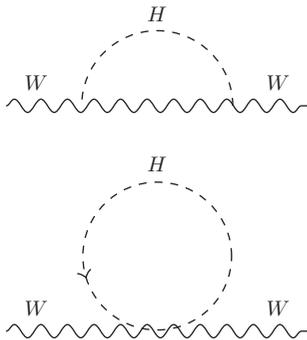


Fig. 7.9 Radiative corrections to the W mass from Higgs loops. Equivalent diagrams also apply to the Z .

The above discussion of the W and Z bosons is valid at lowest order in perturbation theory. There are, however, small but important radiative corrections Δr from higher-order diagrams. The results of these corrections are parameterized by modifying eqn 7.44 to

$$M_W^2 = \frac{\sqrt{2}e^2}{8G_F \sin^2 \theta_W (1 - \Delta r)} \quad (7.46)$$

There are contributions to Δr from fermion loops containing t quarks (in principle, other quarks contribute, but the t quark is dominant because of its much larger mass) as shown in Fig. 7.8 on page 200.

These give a contribution

$$(\Delta r)_{\text{top}} = -\frac{3G_F m_t^2 c_W^2}{8\sqrt{2}\pi^2 s_W^2} \quad (7.47)$$

where we define $s_W = \sin \theta_W$ and $c_W^2 = 1 - s_W^2$. The masses of the W and Z are also affected by Higgs loops (see Fig. 7.9 on page 200). The contribution to Δr is given by

$$(\Delta r)_{\text{Higgs}} = \frac{11G_F M_Z^2 c_W^2}{24\sqrt{2}\pi^2} \ln \frac{m_H^2}{M_Z^2} \quad (7.48)$$

Δr (eqn 7.46) is the sum of the virtual top and Higgs loop corrections to M_W and M_Z as well as the running of the fine structure constant (α) from low energy to the value at M_Z . The effect of the running of α is given by

$$\delta r_0 = 1 - \alpha/\alpha(M_Z)$$

Where α is the value of the fine structure constant at low energy and $\alpha(M_Z)$ is the value at the scale $Q^2 = M_Z^2$. The overall sum of these radiative corrections is given by

$$\Delta r = \delta r_0 + \Delta r_{\text{top}} + \Delta r_{\text{Higgs}}$$

Therefore, precision measurements of M_W , M_Z , and G_F make predictions for the allowed values of the masses of the top quark and the Higgs boson. Note that the contribution from the Higgs depends logarithmically on m_H , whereas the contribution from the top quark scales as m_t^2 . Therefore, even with no knowledge of the Higgs mass other than that imposed by unitarity, precision electroweak measurements including M_W and M_Z predicted the mass of the top quark to be around 170 GeV (see Chapter 8).

7.4.4 The standard model, how good is it?

We now look at a list of some of the difficulties with the weak interaction:

- (1) The original Fermi 4-point theory had a problem, the cross section $\sigma \propto G_F E^2$, where E is the centre-of-mass energy. This is fine at low energy, but at 300 GeV the scattering probability becomes bigger

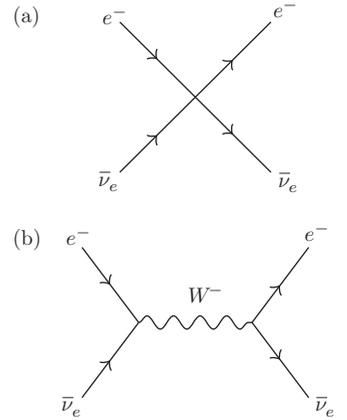


Fig. 7.10 Antineutrino–electron scattering in the original Fermi 4-point theory (a) and including the W intermediate boson (b).

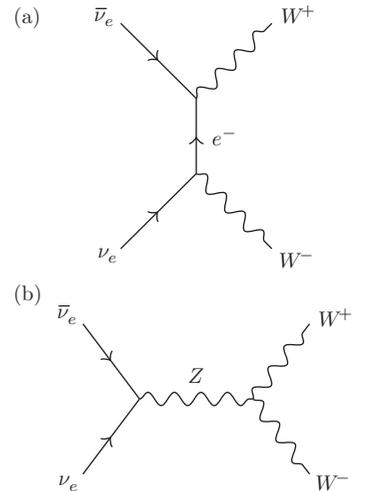


Fig. 7.11 $\nu_e\bar{\nu} \rightarrow W^+W^-$ t -channel, electron exchange (a) and neutral-current, Z exchange (b).

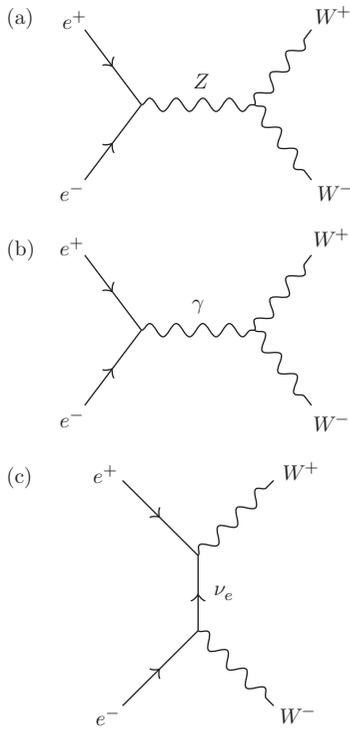


Fig. 7.12 The three diagrams contributing to $e^+e^- \rightarrow W^+W^-$, via Z exchange (a), γ exchange (b), and t -channel neutrino exchange (c).

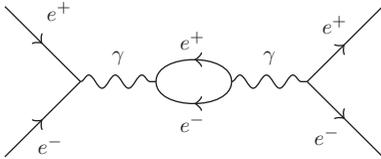


Fig. 7.13 Higher-order Feynman diagram with an e^+e^- loop for the process $e^+e^- \rightarrow e^+e^-$.

than 1. This is known as unitarity violation and is bad news for a theory. By introducing the W boson and moving away from a four-point interaction (Fig. 7.10), a propagator term $1/(M_W^2 - q^2)$ is introduced and the cross section stops increasing.

- (2) Another similar problem occurs even with W bosons in the theory. The process $\nu\bar{\nu} \rightarrow W^+W^-$ in the theory before electroweak unification (Fig. 7.11(a)) is divergent (i.e. becomes very large at high energy). Electroweak theory predicts neutral currents, i.e. diagrams including the Z , so we now have another possible diagram (Fig. 7.11(b)) involving $Z \rightarrow WW$, which cancels the divergence.
- (3) Also, in $e^+e^- \rightarrow W^+W^-$, the two charged-current diagrams shown in Fig. 7.12 need cancellation from the third diagram, which comes from the combined electroweak theory.
- (4) **Renormalization** We have not gone into full detail about calculating Feynman diagrams. Higher-order diagrams contain internal loops. The Feynman rules require that we integrate over the particle momentum in each loop. The diagram in Fig. 7.13 presents a problem in that the integral (which we want to evaluate with an upper limit of infinity) is

$$\int_0^\infty \frac{1}{q^4} q^3 dq = \ln |q| \tag{7.49}$$

which is divergent. The solution, renormalization theory, took a long time to develop and to be shown to work. Eventually, this task was completed by 't Hooft and Veltman, who showed that all locally gauge-invariant theories are renormalizable [130].

- (5) The theory of the EM interaction (QED) is locally gauge-invariant and we can use 't Hooft's theorem to assure ourselves that it is renormalizable. The same is true for QCD.
- (6) For the weak interaction, however, we have a problem. The bosons are *massive* and enter in a different (non-locally gauge-invariant) way to the EM interaction in the Dirac equation. Now 't Hooft's theorem does not apply. This is where the Higgs mechanism for providing particle masses is crucial. The theory is outlined and results on the Higgs from the LHC are given in Chapter 12.

Chapter summary

- All leptonic charged-current weak interactions are described with one coupling constant g_w .
- The weak interaction is determined from experiment to be V-A, i.e. $\gamma_\mu(1 - \gamma_5)$.

- The universality of the weak interaction with coupling g_w also holds for quarks, provided the CKM-rotated states (d' , s' , b') are used for the charge $-\frac{1}{3}$ quarks.
- Electroweak unification is outlined, with charged-current interactions mediated by the W^\pm and neutral-current interactions by the Z^0 and the photon (which remains massless).

Further reading

- Griffiths, D. (2008). *Introduction to Elementary Particles* (2nd revised edn). Wiley-VCH. This gives a very clear introduction to Feynman rules and how to perform the calculations.
- Burcham, W. E. and Jobes, M. (1994). *Nuclear and Particle Physics*. Pearson. This contains more advanced discussion of the electroweak theory and a good introduction to renormalization theory.
- Taylor J. C. (1979). *Gauge Theories of Weak Interactions*. Cambridge University Press. This gives a concise account of the quantum field theory and gauge symmetry underlying electroweak unification.
- Halzen, F. and Martin, A. D. (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley. This is another very good graduate-level textbook explaining the theory of the Standard Model.
- Thompson, M. (2013). *Modern Particle Physics*. Cambridge University Press. This is another recent textbook covering the Standard Model well.

Exercises

- (7.1) Verify the identity $\gamma_\mu \gamma_5 + \gamma_5 \gamma_\mu = 0$. *Hint*: See chapter 6 for the definition of the γ matrices.
- (7.2) Explain why in general the vector coupling constants $c_V^{(f)}$ depend on the weak mixing angle $\sin^2 \theta_W$ but the axial vector coupling constants $c_A^{(f)}$ do not. Why does $c_V^{(f)}$ for the neutrinos not depend on $\sin^2 \theta_W$.
- (7.3) Show that $\bar{u} \gamma_\mu u = \bar{u}_L \gamma_\mu u_L + \bar{u}_R \gamma_\mu u_R$.
Hint: Show that $\bar{u}_R \gamma_\mu u_L = 0$ by adapting eqn 7.25 and then proceed backwards through the steps used to derive the original eqn 7.25; you should find a combination of projection operators that gives zero.
- (7.4) What are the possible decay modes of the τ ?
Given that lifetime of the muon is 2×10^{-6} s, estimate the expected lifetime of the τ ? How might it be measured?

Neglecting density-of-states factors, what is the expected ratio of branching ratios for

$$\frac{\tau^+ \rightarrow K^+ \bar{\nu}_\tau}{\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau} \quad ?$$

Starting with an intense 800 GeV proton beam, how could a high-energy neutrino beam, enriched in ν_τ , be produced? Explain the origin of reducible and irreducible backgrounds of other neutrino flavours.

- (7.5) Draw quark flow diagrams for the following decays and discuss the dependence of the decay rates on the matrix elements of the CKM matrix:
- $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$
 - $K^+ \rightarrow \mu^+ + \nu_\mu$
 - $\pi^+ \rightarrow \mu^+ + \nu_\mu$
 - $D^+ \rightarrow K^- + \pi^+ + \pi^+$
 - $D^+ \rightarrow K^+ + \pi^+ + \pi^-$

The Λ has a mean lifetime of 2.6×10^{-10} s and decays into $p + e^- + \bar{\nu}_e$ with a branching fraction of 8.3×10^{-4} . The $\Lambda_c^+(udc)$ has a mean lifetime of 2.1×10^{-13} s. Estimate the branching fraction of $\Lambda_c^+ \rightarrow \Lambda + e^+ + \nu_e$ and comment on how your result compares with the measured value.

$[m(\Lambda_c^+) = 2.285 \text{ GeV}, \text{BR}(\Lambda_c^+ \rightarrow \Lambda e^+ \nu_e) = (2.1 \pm 0.6)\%.]$

(7.6) Explain how each of the following measurements could be used to determine a single element of the CKM matrix or a combination of elements:

- the ratio of μ -pair production to single- μ production in ν_μ interactions on nuclei;
- the decay rate of $D_s^+ \rightarrow \tau^+ \nu_\tau$;
- the decay rate of $\bar{B}^0 \rightarrow D^{*+} \mu^- \nu_\mu$;
- the region of the μ -momentum spectrum near the kinematic endpoint from $\bar{B} \rightarrow X \mu \nu_\mu$ decays (where X is any hadronic final state).

How are these results affected by the strong interaction?

(7.7) Draw quark flow diagrams for the decays of the charm mesons D^0 and D^+ . In the ‘spectator’ model, the decay rate is only determined by the quark that changes flavour. Explain why this model works quite well for the equivalent beauty mesons but not so well for the charm mesons.

(7.8) Draw two Feynman diagrams for each of the decays $D^0 \rightarrow K^+ \pi^-$ and $D^0 \rightarrow K^- \pi^+$. Estimate the ratio of the two decay rates based on the different CKM matrix elements and explain why the phase-space factors should be the same. Compare your prediction with the measured values in [115] and comment on the origin of any discrepancies.

(7.9) The matrix elements in Z decay are proportional to a factor $c_V - c_A$. Explain why the partial width for a given decay mode is proportional to a factor of $c_V^2 + c_A^2$. The measured partial widths of the Z in decays to hadrons and to electrons are $\Gamma_Z^{\text{had}} = 1744 \pm 2.0 \text{ MeV}$ and $\Gamma_Z^{e^e} = 83.92 \pm 0.12 \text{ MeV}$. Determine the compatibility of the ratio of these two measurements with the SM and a value of $\sin^2 \theta_W = 0.231$ and explain any discrepancies between your prediction and the experimental value.

8.1 Neutrinos	205
8.2 Charged currents	206
8.3 Neutral currents	208
8.4 Physics at e^+e^- colliders	208
8.5 W and Z physics at hadron colliders	224
8.6 Top-quark physics	229
8.7 Summary	232
Chapter summary	233
Further reading	233
Exercises	233

Experimental tests of electroweak theory

Chapter 7 covered the basic ideas of weak-interaction theory and the unification of the electromagnetic and weak interactions. The weak charged-current interaction is expressed using a unique coupling constant g_w for all leptonic vertices: (ν_e, e) , (ν_μ, μ) , (ν_τ, τ) . The coupling constant is also valid for charged-current interactions involving quarks using the weak quark eigenstates, (u, d') , (c, s') , (t, b') , which are a rotation of the flavour eigenstates. We showed how the absence of flavour-changing weak neutral currents is explained using the Glashow, Iliopoulos, and Maiani (GIM) mechanism. We gave an outline of how the weak and electromagnetic interactions are unified into a single consistent theory. The resulting electroweak theory includes only one additional parameter, $\sin \theta_W$, and predicts all the features of the weak neutral current.

In this chapter, we review some of the experimental evidence that underpins the electroweak theory. We start with some of the key neutrino experiments, then look at the discovery of neutral currents, as this was the first step towards a unified electroweak theory. The key prediction of the theory was the existence of the massive W and Z bosons, with quite precise estimates of their masses. Their discovery with masses in the predicted range was a triumph! Electroweak theory has now been probed to much higher precision by many experiments, particularly at LEP and the Tevatron. Many details have since been filled in, but the basic structure remains unchanged.

8.1 Neutrinos

When Pauli postulated the existence of neutrinos, he was afraid that the cross sections were so small that they would never be measurable. The experimental discovery of neutrinos was made possible by the intense flux of antineutrinos from nuclear reactors.¹ The reaction studied was $\bar{\nu}_e p \rightarrow e^+ n$, using water as the target. Each e^+ annihilated with an e^- to produce two photons. The photons were detected using tanks of liquid scintillator viewed by photomultipliers. To reduce the backgrounds, cadmium chloride was added to the water. This allowed neutrons to be captured by $n^{108}\text{Cd} \rightarrow ^{109}\text{Cd} \gamma$. The neutron capture happened a few microseconds after the first reaction. Therefore, a clean signal for the reaction was a flash of light followed by a delayed coincidence.²

¹Later, high-energy proton accelerators were used to create neutrino beams at much higher energies (see Chapter 11).

²The proof that the events were from the reactor as opposed to backgrounds like cosmic rays was provided by running with the reactor off.

If detecting neutrinos was challenging, how could one measure the helicity of a neutrino? The answer came in a brilliant experiment [79]. The key idea was to transfer the helicity of the neutrino to a photon, which could be measured relatively easily. The source of the neutrinos was electron capture $\text{Eu } e^- \rightarrow \text{Sm}^* \nu_e$, with the subsequent decay $\text{Sm}^* \rightarrow \text{Sm } \gamma$. As the energy is shared between the Sm and the γ , in general the γ does not have sufficient energy to be absorbed by another Sm nucleus. However, if the Sm^* decays with the γ travelling in the direction of the Sm^* , then it will have sufficient energy to make a resonant scatter off Sm (see Exercise 8.1). The photon must have the same helicity as the neutrino (see Exercise 8.1). The experimental apparatus is sketched in Fig. 8.1. Photons of the correct energy will be resonantly scattered by the ring of Sm_2O_3 and then detected by the NaI(Tl) scintillator coupled to a photomultiplier. The rate is measured with two polarities of the magnetic field. The scattering cross section for γSm is greater if the spin of the photon is anti-aligned with that of the iron than if it is aligned.³ The polarization of the photons could thus be determined and hence the helicity of the neutrino also. It was found to be consistent with -1 , as expected.

The concept of lepton number was invented to explain the absence of decays $\mu \rightarrow e\gamma$. Neutrinos must also carry lepton number. The experimental demonstration [69] that ν_e are distinct from ν_μ came from an experiment in which a ν_μ beam was fired at a 5000 ton steel wall⁴ to absorb all particles other than neutrinos. The neutrinos then interacted in aluminium plates and the resulting charged particles were detected by spark chambers. Muons could be separated from electrons because of their longer range. The observation of muons coupled with the absence of electrons showed that ν_μ were distinct from ν_e .

When the τ lepton was discovered, it was assumed that there would be an associated neutrino, ν_τ . The experimental confirmation of the ν_τ was made by the DONUT Collaboration [92]. Producing a ν_τ beam is quite a challenge. The first step was to direct an intense beam of 800 GeV protons at a tungsten target. The forward-going interaction products then entered a magnetic field, which swept charged particles aside, greatly enhancing the neutrino content of the beam, including ν_τ (from the decays of charm mesons, such as the D_s). As for other neutrinos, to identify a ν_τ it must first interact to produce a charged τ , which can then be detected through its charged-current interaction, $\nu_\tau X \rightarrow \tau Y$. The short lifetime of the τ leads to tracks with ‘kinks’ near the primary vertex. These kinks were identified in an emulsion chamber, but in order to determine which volume of the emulsion to measure, a magnetic spectrometer was used to identify candidate τ events. Four such events were found, significantly above the background of 0.34 events.

³The electron spins responsible for ferromagnetism are aligned antiparallel to the \mathbf{B} field.

⁴The steel plates used came from an old battleship.

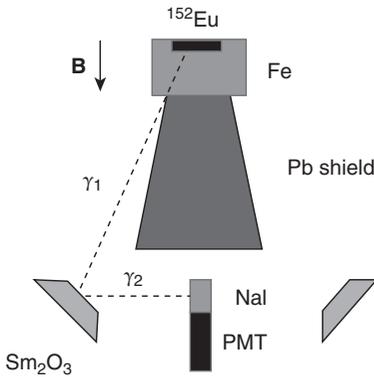


Fig. 8.1 Schematic of the experiment to measure the neutrino helicity.

8.2 Charged currents

The theory of charged currents developed in Chapter 7 is based on a $V-A$ structure. This is parity-violating, and the first clear experimental

observation of parity violation was in the decay ${}^{60}\text{Co} (J^P = 5^+) \rightarrow {}^{60}\text{Ni}^* + e^- + \nu_e$. The ${}^{60}\text{Co}$ nuclei were aligned by an external magnetic field.⁵ The rate of emission of electrons was found to be consistent with an angular distribution of the form $1 - (v/c) \cos \theta$, where θ is the angle between the electron 3-momentum and the magnetic field direction, as predicted by a V–A interaction. A more recent and more direct demonstration of parity violation is given by the angular distribution of leptons from W decays (see Section 8.5.1). Very strong evidence for the V–A theory also comes from the ratio of branching ratios $\text{BR}(\pi^+ \rightarrow e^+ \nu_e) / \text{BR}(\pi^+ \rightarrow \mu^+ \nu_\mu)$; see Exercise 6.4. The cleanest probe of the V–A theory comes from muon decays $\mu \rightarrow e \bar{\nu}_e \nu_\mu$. Using intense muon beams (from pion decays), very precise measurements of the electron spectrum from stopped muons agree with the V–A theory and place stringent limits on contributions from other interactions.

⁵This required adiabatic demagnetization to achieve sufficiently low temperatures, 0.01 K.

8.2.1 Measurements of CKM matrix elements

In Chapter 7, we outlined how the theory of charged-current weak interactions of leptons could be extended to quarks with a universal coupling strength. This required the CKM matrix to allow for the rotation between the weak and mass eigenstates of the quarks. The theory gives no predictions for the values of these matrix elements, so they have to be determined experimentally. However, as multiple experiments can be performed to determine the same element of the CKM matrix, powerful consistency checks of the theory can be performed. The CKM matrix should be unitary, so this gives additional constraints on the theory. This aspect will be developed in Chapter 10. We give a very brief outline here of some of the methods used to measure the CKM matrix elements.⁶

⁶See the review from the Particle Data Group in Further Reading for a comprehensive discussion.

Measurement of CKM matrix elements proceeds by measuring many different processes; for example V_{ud} is measured by comparing hadronic beta decays with the decay of the muon, and V_{us} is measured by comparing the decay $K \rightarrow \pi e \nu$ with a non-strange decay.

The ratio of V_{ub}/V_{cb} can be determined from the muon spectrum in the decays $b \rightarrow c \mu \bar{\nu}_\mu$ and $b \rightarrow u \mu \bar{\nu}_\mu$. As the u quark is much lighter than the c quark, the spectrum of muons from $b \rightarrow u$ decays extend beyond the end of those from $b \rightarrow c$. This enables a clean sample of muons from $b \rightarrow u$ to be identified despite the fact that the ratio $V_{ub}/V_{cb} \ll 1$.

The ratio V_{cd}/V_{ud} can be measured by observing the rate of dimuon to single-muon production in charged-current neutrino interactions on hadrons. If the neutrinos are above threshold, they can produce either charm or up quarks (see Fig. 8.2). A known fraction of the events with a charm quark will result in the semimuonic decay of the charm quark ($c \rightarrow \mu X$) and hence result in events with two muons, whereas the events in which an up quark was produced will result in events with single muons. As the Feynman diagrams are the same for the two processes, the difference in the rates is simply given by the ratio of the CKM elements V_{cd}/V_{ud} once $\text{BR}(c \rightarrow \mu X)$ has been accounted for.⁷

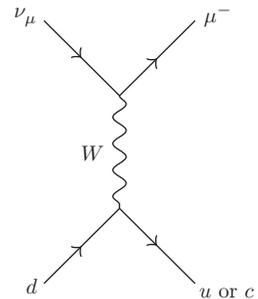


Fig. 8.2 Feynman diagram for $\nu_\mu d \rightarrow \mu c(u)$.

⁷We have assumed that the energy is far above threshold.

8.3 Neutral currents

The neutral current was first discovered in the Gargamelle bubble chamber at CERN in the reaction $\bar{\nu}_\mu e \rightarrow \bar{\nu}_\mu e$, elastic scattering of $\bar{\nu}_\mu$ off atomic electrons. The scattering reaction of either ν_μ or $\bar{\nu}_\mu$ off electrons is an unambiguous signal of a neutral current. Scattering of either ν_e or $\bar{\nu}_e$ is ambiguous since there is a charged-current diagram for each of these reactions (see Fig. 8.3).

About 100 events were observed in total and, by measuring the cross section, a value of $\sin^2 \theta_W = 0.24 \pm 0.04$ was obtained. This was the first success for the unified electroweak theory and it enabled the prediction of the masses of the W and Z bosons (see Section 8.5.1).

Subsequent studies of weak neutral currents with neutrino beams used electronic detectors that allowed the accumulation of much larger numbers of events. For example, the CHARM2 experiment used the neutral-current reactions $\nu_\mu e^- \rightarrow \nu_\mu e^-$ and $\bar{\nu}_\mu e^- \rightarrow \bar{\nu}_\mu e^-$. The ratio of the cross sections at the same energy is given by (see Exercise 8.2)

$$R = 3 \frac{1 - 4 \sin^2 \theta_W + \frac{16}{3} \sin^4 \theta_W}{1 - 4 \sin^2 \theta_W + 16 \sin^4 \theta_W} \quad (8.1)$$

An intense beam of neutrinos was used and particular care was taken to reduce backgrounds. For example, consider a neutral-current interaction on a nucleus, $\nu_\mu N \rightarrow \nu_\mu \pi^0 X$. The photons from the π^0 decay may pair-convert ($\gamma \rightarrow e^+ e^-$) in the detector material, potentially faking the single-electron signature. However, the electron energy (E_e) and angle (θ_e) with respect to the neutrino beam are limited by (see Exercise 8.3)

$$E_e \theta_e^2 < 2m_e \quad (8.2)$$

Hence the quantity $E_e \theta_e^2$ will be peaked at small values on top of a continuous background. This requires a neutrino detector with very good angular and energy resolution. The target was made from glass since this contains elements of relatively low atomic number and hence minimizes multiple scattering, which limits the angular resolution (see Chapter 4). The final result from the CHARM2 [133] experiment was $\sin^2 \theta_W = 0.2324 \pm 0.0083$.

8.4 Physics at $e^+ e^-$ colliders

$e^+ e^-$ machines are the place of choice to study the Z^0 . The reasons are as follows:

- e^+ and e^- are point-like and the centre-of-mass (CMS) energy is fixed.
- Unlike a hadron collider, there is no underlying event (see below) from the beam-particle spectator quarks not involved in the hard collision.
- Positrons are much easier to produce than antiprotons.

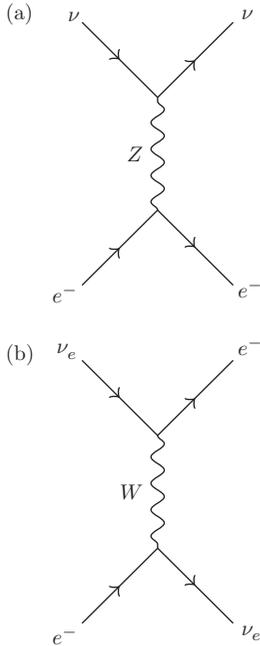


Fig. 8.3 (a) $\nu e^- \rightarrow \nu e^-$, for any neutrino flavour, via a neutral current, with Z exchange. (b) $\bar{\nu}_\mu e \rightarrow \bar{\nu}_\mu e$, for an electron neutrino, via a charged current, with W exchange.

Because of the finite size of hadrons, an ‘underlying event’ in a hadron–hadron collider is a superposition of multiple quark or gluon collisions occurring in the same hadron–hadron collision as the ‘hard scatter’ of interest between hadron constituents. Usually, the underlying event consists of particles produced at small angles to the beam axis.

The big disadvantage is that electrons and positrons in circular colliders lose energy through synchrotron radiation.⁸ All charged particles, when accelerated, will radiate energy at a rate $\propto 1/m^4$. So, for electrons and protons in circular colliders of the same radius, the ratio of energy loss is $(m_p/m_e)^4 \sim 10^{13}$. The alternative of using a linear e^+e^- collider is discussed in Chapter 13.

A chain of accelerators⁹ (see Table 8.1) was used to produce the e^+e^- beams and increase their energies up to the target 45 GeV per beam in the LEP ring (Fig. 8.4). The chain started with linear accelerators (LINAC) and then progressed through three circular synchrotron machines (PS, SPS, and LEP). Filling took about one hour, with beams accumulating in LEP at 20 GeV (two cycles every 14.4 s). The counter-circulating LEP beams were then accelerated from 20 GeV to 45 GeV and left ‘coasting’ (i.e. there was no further acceleration—the radiofrequency cavities were used just to replace energy lost through synchrotron radiation) for about 8 hours, during which time the e^+e^- beams collided at the positions of the four detectors at LEP: ALEPH, DELPHI, L3, and OPAL.

⁸More details are given in Section 3.2.2.

⁹See Chapter 3 for an explanation of the need for a chain of accelerators between the source and the high-energy ring.

LINAC	$e^- \rightarrow 200$ MeV e^+ produced in converter (EM shower)
LINAC	$e^+, e^- \rightarrow 600$ MeV
PS	$e^\pm \rightarrow 3.5$ GeV
SPS	$e^\pm \rightarrow 20$ GeV
LEP	$e^\pm \rightarrow 45$ GeV

8.4.1 Detailed look at the detectors

The detectors were of a roughly cylindrical shape providing almost ‘ 4π ’ (steradian) coverage; i.e. they were sensitive to particles going in any direction from the interaction point (with no cracks or holes except for the beam pipe through the centre). This was very important for some of the analysis techniques we are going to discuss later. The OPAL detector is shown in Fig. 8.5 as a typical example. The LEP detectors were general-purpose collider detectors as described in Chapter 4. There

Table 8.1 Chain of accelerators for LEP.

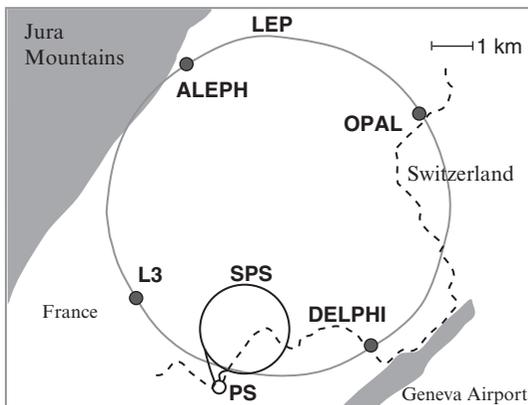


Fig. 8.4 Map of the accelerator complex at CERN and the four LEP detectors. From [17].

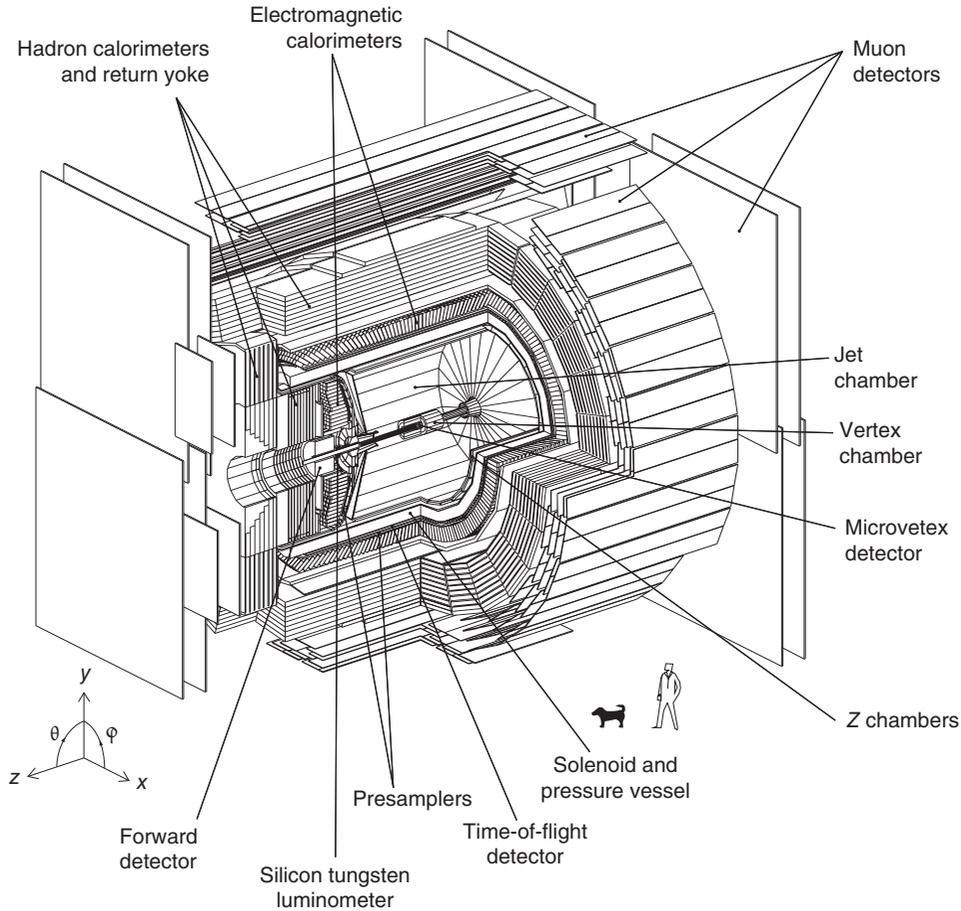


Fig. 8.5 The OPAL detector. From [17].

were interesting differences between the detectors because of the different optimization strategies employed. For example, the L3 had a very high-resolution electromagnetic calorimeter based on BGO (bismuth germanium oxide). BGO is a dense crystal with a good scintillation yield and was used as a homogeneous calorimeter. However, the smaller size and lower value of the magnetic field meant that the charged-track resolution was not as good as for the other LEP experiments.¹⁰ Crucial parts of all four of the LEP detectors were the vertex detectors, which used silicon as the active material. These provided high-resolution tracking close to the beam pipe, which enabled the identification of jets from b quarks using the relatively long lifetime of B mesons (see Chapter 4).

We will now summarize the different types of events seen at LEP when running at the Z^0 resonance as a review of what particles do as they pass through material. What the detectors ‘see’ is visualized using event displays, examples of which are shown in Fig. 8.6. Event displays are

¹⁰This is another demonstration that there is no perfect detector design.

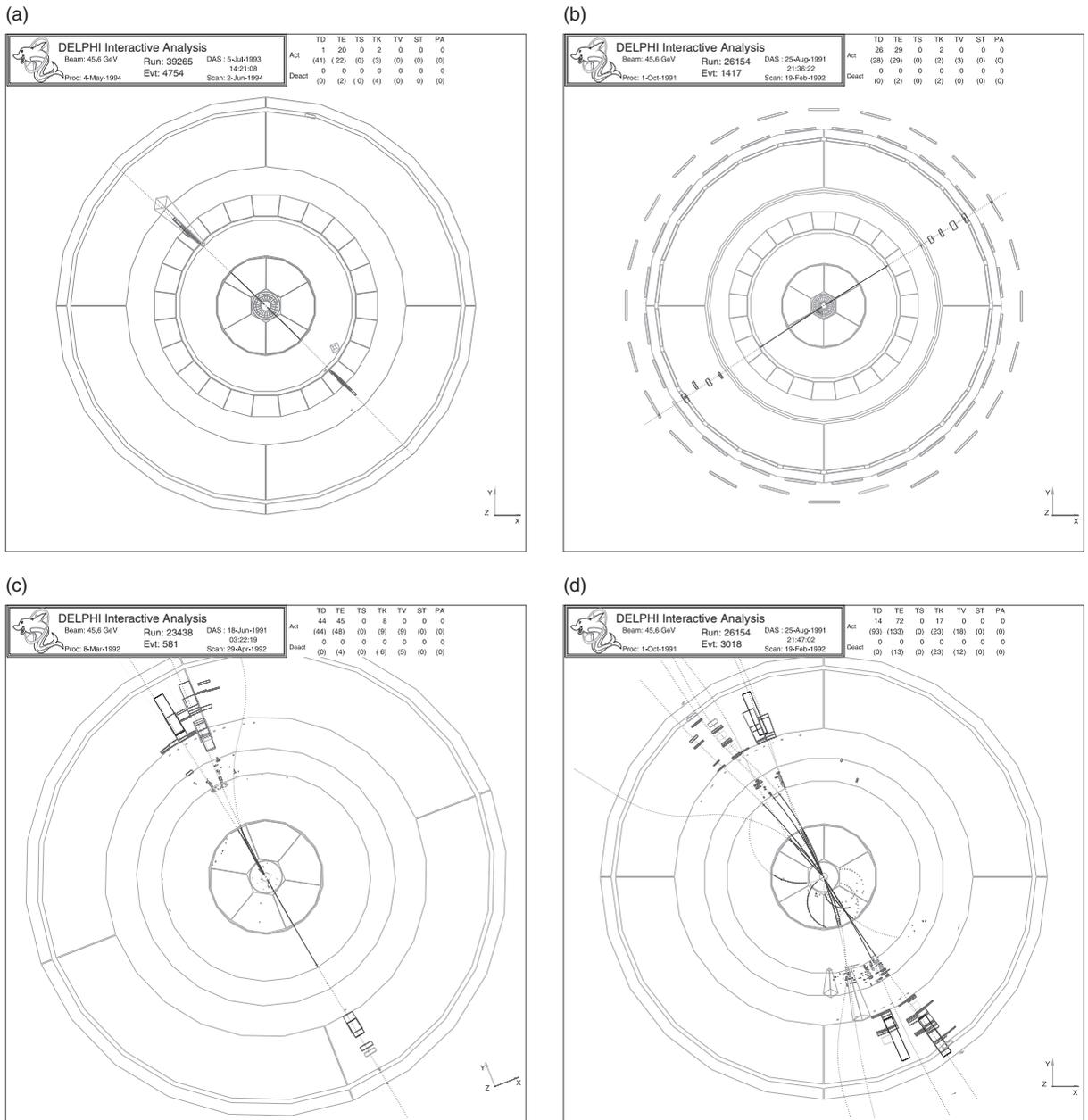


Fig. 8.6 Four event types at LEP from the DELPHI experiment (a) e^+e^- ; (b) $\mu^+\mu^-$; (c) $\tau^+\tau^-$; (d) quark-antiquark pair.

important tools for checking that the detectors are functioning correctly, for pedagogic purposes, and for examining unusual events.

- $Z \rightarrow e^+e^-$: Two back-to-back showers in electromagnetic calorimeter with tracks leading towards them. Nothing in hadron calorimeter or muon detectors.
- $Z \rightarrow \mu^+\mu^-$: Two back-to-back tracks leading all the way out to the muon detectors.
- $Z \rightarrow \tau^+\tau^-$: A variety of topologies that have low multiplicities and missing momentum (from the neutrino(s)).
- $Z \rightarrow \nu\bar{\nu}$: Nothing (see Exercise 8.4).
- $Z \rightarrow$ 2-jets, $Z \rightarrow q\bar{q}$: The quarks form ‘jets’—groups of mesons and baryons collimated along the initial quark direction. Generally, it is not possible to tell what flavour the original quark had.
- **b-jets**: For b (and c) quarks, using the silicon vertex detectors, it is possible to detect whether the particles all extrapolate back to the primary vertex or whether there is a secondary vertex where the b quark decayed.
- $Z \rightarrow$ 3-jets, $Z \rightarrow q\bar{q}$: With a gluon ‘bremsstrahlung’ emitted from one of the quarks, also materializing as a jet.

The event categories can be distinguished—in particular e^+e^- and $\mu^+\mu^-$ —by their distinctive two-particle topologies. A plot of the total invariant mass of all the particles in the event versus the number of charged particles is shown in Fig. 8.7. The $\tau^+\tau^-$ events can be distinguished from 2-jet events with these variables.

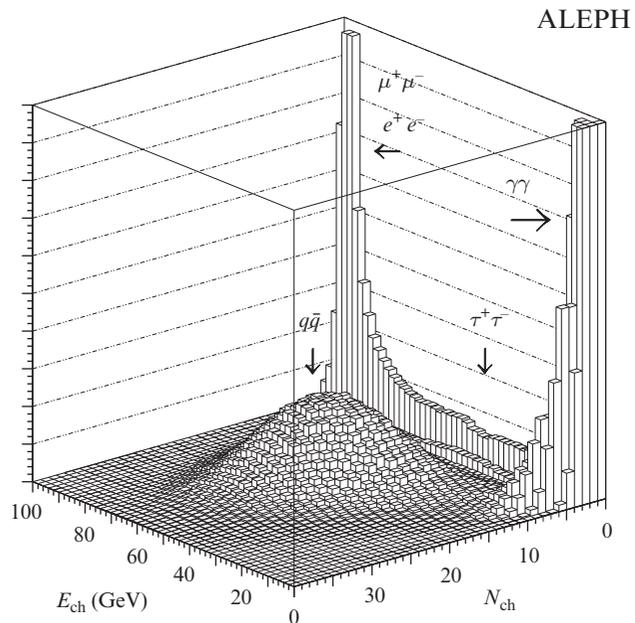


Fig. 8.7 Main cuts used to separate the different classes of e^+e^- events. From [43].

8.4.2 Aspects of a physics analysis

What is written here is valid for the analysis of any high-energy physics data, but is described in the context of a LEP experiment. An analysis generally goes along the following lines: the experimenters first choose selection criteria (**cuts**), which select the desired events, the **signal**, such as requiring n tracks that look like muons, or tracks above a certain energy, etc. This may involve reconstructing the mass of a combination of particles from the measured 4-momenta of the tracks and showers.¹¹ Quite often there is **background** in the sample selected—where a different physical process produces events that pass the same cuts. Some backgrounds are ‘irreducible’ in that they produce the same final-state particles as the signal; other backgrounds are ‘reducible’ in that they could be reduced by tighter cuts or with a better detector.

Calibration concerns subtracting pedestals¹² and measuring the gains and linearities of each channel, particularly for calorimeters. If the calibration is not done correctly, then the energy resolution of the calorimeter suffers. The **acceptance** for the process we are studying has to be calculated. Acceptance is the probability of the particles in an event hitting a certain part of the detector and having a certain minimum energy or momentum.¹³ **Accidentals** or pile-up concerns the problem when two events occur at the same time (within detectable resolution) and the resulting combination gets into the data sample. Accidentals can also be a problem if a preceding event causes electronics to momentarily become inactive before being ready to measure a new pulse, or if the preceding event causes a movement in the pedestals that causes the energy to be measured slightly wrongly. Accidentals were not a big problem in LEP experiments, because the rate of events was low.

8.4.3 Monte Carlo simulation

An essential tool for any analysis of data from a large particle physics detector is the Monte Carlo simulation computer program.¹⁴ The idea is to produce simulated events, using a random number generator at each point where a choice in what happens must be made. Both the physics process and the detector response are simulated.

Figure 8.8 shows a block diagram of the steps involved from the choice of physics channel to be simulated through the generation of the detector response, then the digitization of the Monte Carlo ‘data’ in an identical format to that produced by the various detector components that make up the complete detector. The Monte Carlo data are then run through the complete ‘real’ data reconstruction chain. They are then available for analysis by the same analysis codes that are used for the real data—the only difference is that one knows what physics process was used to generate the events.

First is the ‘physics simulator’: physics events are simulated starting from theoretical matrix elements for the basic quark and gluon scattering processes that could occur in a hard scatter. The output will be

¹¹Sometimes we assume that the combination has a mass $m = 0$.

¹²Even if there is no genuine signal in a detector channel, the readout will still deliver a non-zero value. This value is called the ‘pedestal’, from the shape of the distribution. The pedestal value must be measured and subtracted from genuine signals.

¹³Usually for a 4π LEP detector, it was almost 100% however, this could have been reduced if cuts were made around any dead channels.

¹⁴Invented by Ulam and Metropolis at Los Alamos in the 1940s and named after the the city in Monaco where there is a big casino.

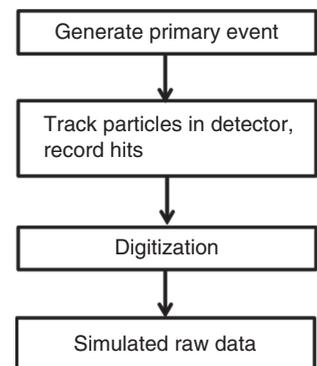


Fig. 8.8 Monte Carlo flow diagram.

the type and 4-momentum of quarks, gluons, and leptons produced. In some cases, for example $e^+e^- \rightarrow \tau^+\tau^-$ and electromagnetic processes, the matrix elements can be calculated. The procedure is similar for hard quark and gluon scattering. The underlying event is produced from so-called phenomenological models that describe the observed properties of the small-angle scattering involved. The physics events can also be chosen to be all of one type if one is trying to work out the best way to select events from a particular physics source (e.g. top-quark production). Sometimes, it is useful to generate only a single type of particle, for example if one needs to know what type of signals it will produce in the different parts of the detector.

Next is the ‘detector simulator’: this is usually the most difficult part to produce and requires a thorough understanding of how the detector components work.¹⁵ Random number simulation is used to decide how each particle proceeds as it passes through various types of material in the detector. For some particles, for example electromagnetic particles (e^\pm, γ), the response of most materials is well documented, and well-honed computer codes exist that can be modified to whatever geometry is required. For hadronic particles, things are more complicated, first because most of these particles are unstable and will decay either within the beam pipe or within the detector layers nearest to the beam. The decays have to be simulated and the resulting products (mostly pions, together with some kaons and nucleons) followed through the detector layers until they are absorbed. The exceptions are neutrinos (which will leave no signal in a typical collider detector) and high-energy muons, which will penetrate through the main detector and surrounding magnet and so require special and very large-area but fairly simple charged-particle detectors covering the outside of the main detector. The final step is to collect the simulated signals, format them as though they were real data, and add the ‘book keeping’ records. These steps are summarized in the remaining three boxes in the flow diagram in Fig. 8.8.

¹⁵Detector simulation is very computer intensive and this is another aspect of modern high-energy physics where GRID computing is essential.

Example

As an example, in generating a simulated $e^+e^- \rightarrow Z^0 \rightarrow \tau^+\tau^-$, the choices could be as follows:

- (1) Pick the direction along which the τ^+ travels (using the appropriate angular distribution) and decays, then use momentum conservation to work out the direction of the τ^- .
- (2) Choose the decay time for each τ by picking from an exponential probability distribution e^{-t/τ_τ} , where τ_τ is the mean decay time of the τ lepton, and work out where it decays.
- (3) Choose from a table of measured branching ratios what daughter particles each τ will decay into.
- (4) Follow the decay products until they decay, are absorbed in the detector material, or exit (e.g. neutrinos).

The technique works well and, by generating lots of simulated events, we are effectively performing a numerical integral over all the possible outcomes of what might happen by randomly sampling the integrand. The Monte Carlo technique is often used to estimate trigger efficiencies, acceptances, and accidental effects in an analysis.

However, there are always systematic uncertainties associated with Monte Carlo calculations, particularly at hadron colliders. Therefore, wherever possible, the calculations should be done in a ‘data-driven’ way that does not rely so heavily on Monte Carlo calculations. This approach will be described in Chapter 13.

Multivariate analysis methods

The above description of an analysis is based on the simplest ‘cut-and-count’ approach in which one makes a series of selections on a sequence of variables and counts the number of events that pass all the selections. This approach is clearly not optimal if there are correlations between the variables that are different for the signal and background processes. Consider a toy example in which one is using two variables x_1 and x_2 to discriminate between signal and background. A cut-and-count analysis would select a rectangular region in (x_1, x_2) space (see Fig. 8.9). However, a more powerful discrimination between the signal and the background might be obtained by selecting a ‘triangular’ region (see Fig. 8.9). In a typical analysis, we have to deal with many variables, so the optimization of the selection is non-trivial. Powerful statistical techniques like neural networks and boosted decision trees are used (see Behnke et al. in Further Reading).

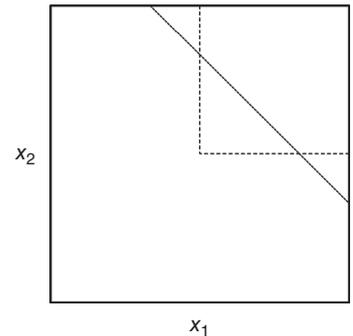


Fig. 8.9 Event selection in the (x_1, x_2) plane.

8.4.4 Physics at LEP

LEP operation was in two phases: for LEP1, the CMS energy was close to the mass of the Z^0 . Similar physics was studied at the SLAC Linear Collider (SLC), where the luminosity was much lower than at LEP, although SLC had the advantage of being able to produce longitudinally polarized electrons. In LEP2, the energy was increased to above the threshold for W^+W^- pair production.

We will review selected LEP1 physics in this section and discuss LEP2 physics in Section 8.4.11. LEP produced a huge number of Z^0 s, $\sim 4.5 \times 10^6$ per detector. Z^0 s decay into almost every type of particle we know, so many things can be studied. Examples include the following:

- $b\bar{b}$: Lifetime of b quark
 $B^0-\bar{B}^0$ mixing
 B^0 CP violation (this is done better at BaBar, Belle, CDF, and LHC)
- $\tau^+\tau^-$: Branching ratios (pre LEP, there was a crisis because $\sum \text{BR} > 100\%$!)
 Decay parameters (information on spins etc. gives information on W and Z currents)

Jets:	QCD tests and precision α_s measurements
Higgs, SUSY:	Searches for production and decay of these particles and other similar exotica
Others:	Measurements of individual particle branching ratios

There were also a number of individual measurements of high importance that we discuss here in more detail: the mass of the Z^0 boson, the number of neutrinos, and production cross section and decay parameters of the Z^0 , which are used to obtain the couplings $c_V^{(f)}$ and $c_A^{(f)}$ to compare with the predictions of the electroweak theory.

The principle of all the measurements at LEP was to take runs at a variety of different beam energies around the Z^0 peak. The experiments recorded everything whenever there was a trigger (\equiv an event), and these were reconstructed later (offline). They were then classified as ee , $\mu\mu$, $\tau\tau$, $q\bar{q}$, or luminosity Bhabha events (see Exercise 3.7). It was then possible to measure the cross section as a function of CMS energy \sqrt{s} and partial cross sections of various types (e.g. according to the decay mode, or which direction the particles went).

8.4.5 The Z line shape

The cross section as a function of \sqrt{s} (at LEP, twice the beam energy) displays a clean peak at the Z resonance. The cross section as a function of s for $e^+e^- \rightarrow Z^0 \rightarrow f\bar{f}$ (where f is one of e , μ , τ , or a quark) is given by

$$\sigma_f(s) = \sigma_f^0(s) \frac{s\Gamma_Z^2}{(s - M_Z^2)^2 + m_Z^2\Gamma_Z^2} \otimes (\text{QEDcorr.}) \otimes (\text{QCDcorr.}) \quad (8.3)$$

where

$$\sigma_f^0(s) = \frac{12\pi}{M_Z^2} \frac{\Gamma_e\Gamma_f}{\Gamma_Z^2} \quad (8.4)$$

and the convolutions take account of the higher-order QED and QCD corrections. Using the electroweak theory from Chapter 7 to compute Γ_f gives

$$\Gamma_f = \frac{G_F\sqrt{2}M_Z^3}{12\pi} \left[\left(c_V^{(f)} \right)^2 + \left(c_A^{(f)} \right)^2 \right] N_{\text{colours}} \otimes (\text{QCDcorr.}) \quad (8.5)$$

Equations 8.3 and 8.4 when put together give the usual relativistic Breit–Wigner formula.

Recall that c_V and c_A are defined for the left- and right-handed parts of a particle separately and vary depending on what type of fermion f we have (see Table 7.6 and the discussion on page 199): $c_V^{(f)} = I_3(f) - 2Q(f)\sin^2\theta_W$ and $c_A^{(f)} = I_3(f)$.

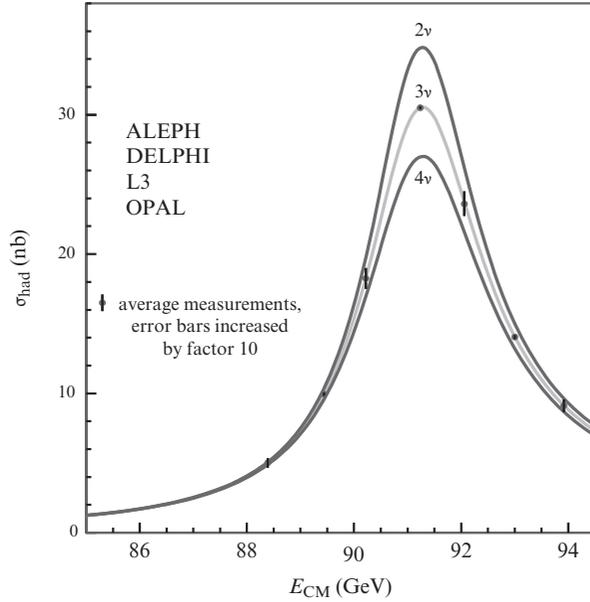


Fig. 8.10 The $e^+e^- \rightarrow Z^0$ production cross section as a function of CMS energy. From [17].

8.4.6 Z mass

The mass of the Z is obtained in principle simply by taking the curve of the cross section as a function of CMS energy (as shown in Fig. 8.10) and fitting the expression given in eqns (8.3)–(8.5) to find the best value for M_Z . For reasons we will come to, it is important to measure M_Z very accurately. In practice, there are various complications to take care of, including the QED and QCD corrections indicated in the formulae. Another important point is to know exactly what the beam energy is, which we describe later.

8.4.7 The Z width; number of neutrinos

The Z^0 can decay into a pair of neutrinos: $Z^0 \rightarrow \nu\bar{\nu}$. How many generations of leptons are there? We know of three: e , μ , and τ . Provided the mass of the associated neutrino is less than half the Z^0 mass, there is a way to detect if there are any more. The total width of the Z^0 , Γ_Z , is made up of the sum of the partial widths of all its decay modes:

$$\Gamma_Z = \Gamma_{\text{had}} + \Gamma_{ee} + \Gamma_{\mu\mu} + \Gamma_{\tau\tau} + N_\nu\Gamma_\nu \quad (8.6)$$

$$= \Gamma_{\text{had}} + 3\Gamma_{\ell\bar{\ell}} + N_\nu\Gamma_\nu \quad (8.7)$$

where lepton universality has been assumed for the second step. The fit to the experimental data is sensitive to both the width of the peak and its height. The partial widths can be predicted from the electroweak model (eqn 8.5) and the branching ratios can be used as a check. Therefore, by

¹⁶Again by fitting the shape of the cross section as a function of beam energy using eqns 8.3–8.5 with the partial widths held fixed at their predicted values from the standard model.

measuring the total width Γ_Z ,¹⁶ the number of neutrinos, N_ν , can be measured. However, since

$$\sigma_0 \propto \frac{1}{\Gamma_Z^2} \quad (8.8)$$

(eqn 8.4), the greatest sensitivity is obtained by simply measuring the cross section at the very top of the peak. The result is

$$N_\nu = 2.9841 \pm 0.0083 \quad (8.9)$$

8.4.8 LEP beam energy measurement

The measurement of the Z mass was a very careful study and involved measuring the beam energy accurately. How do you measure the beam energy? A technique called resonant depolarization involving the anomalous magnetic moment of the electron ($g - 2$) was employed (recall the experiment to measure accurately the anomalous magnetic moment of the muon as one of the most stringent tests of QED). Let us approximate LEP by a circle immersed in a uniform vertical \mathbf{B} field (see Fig. 8.11), which provides the bending. Then

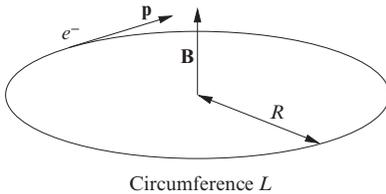


Fig. 8.11 Bending field in the LEP accelerator.

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = -e\mathbf{v} \times \mathbf{B}, \quad |\mathbf{p}| = eBR = \frac{e}{2\pi}BL \quad (8.10)$$

Also, the orbital angular frequency is $\omega_c = eB/\gamma m_e$. Electrons ‘naturally’ become polarized over about 5 hours while going around in LEP. The polarization can be measured with backscattered light. The spin precession angular frequency is given by

$$\omega_s = \frac{eB}{\gamma m_e} \left[1 + \gamma \left(\frac{g-2}{2} \right) \right] \quad (8.11)$$

so we can compute the number of precessions per turn in LEP, ν_s :

$$\nu_s = \frac{\omega_s - \omega_c}{\omega_c} = \gamma \left(\frac{g-2}{2} \right) = \frac{E_{\text{beam}}}{m_e} \left(\frac{g-2}{2} \right) \quad (8.12)$$

The value of $(g-2)/2$ is known to an accuracy of 4×10^{-9} and the electron mass m_e to a precision of 3×10^{-7} , so if we can measure ν_s , we get the energy of the beam.

The technique for measuring ν_s (resonant depolarization) proceeds by adding a small magnet with a field in the x direction (horizontal, transverse to the beam) that varies as $\sin \nu t$. When $\nu = \nu_s$, the contribution will accumulate each turn and cause the beam to depolarize (as measured in the backscattered light). The precision obtained in the beam energy at LEP was 2 MeV (out of 45 GeV). It involved understanding various effects, including the tidal pull of the moon (which changes L slightly), movements in the water table, and even ground currents caused when the fast TGV trains to Paris passed by.¹⁷

¹⁷This was discovered on a day when there was a rail strike!

8.4.9 Cross sections and forward–backward asymmetries at the Z

The couplings c_V and c_A for each fermion type can be extracted from measurements of $d\sigma/d\Omega$ and the forward–backward asymmetry A_{FB} , defined as

$$A_{FB} = \frac{N_F - N_B}{N_F + N_B} \quad (8.13)$$

where N_F and N_B are the numbers of events with $\theta < 90^\circ$ (forward) and $\theta > 90^\circ$ (backward), respectively.¹⁸

Starting from eqn. (7.40), we can show that (see Exercise 8.7)

$$\frac{d\sigma}{d\Omega} = \frac{G_F^2 M_Z^4}{32\pi^2 \Gamma_Z^2} [A(1 + \cos^2 \theta) + B \cos \theta] \quad (8.14)$$

where

$$A = \left[\left(c_V^{(e)} \right)^2 + \left(c_A^{(e)} \right)^2 \right] \left[\left(c_V^{(f)} \right)^2 + \left(c_A^{(f)} \right)^2 \right], \quad B = 8c_V^{(e)} c_A^{(e)} c_V^{(f)} c_A^{(f)}$$

Integrating eqn (8.14), we find

$$\sigma(e^+e^- \rightarrow f\bar{f}) \propto \left[\left(c_V^{(e)} \right)^2 + \left(c_A^{(e)} \right)^2 \right] \left[\left(c_V^{(f)} \right)^2 + \left(c_A^{(f)} \right)^2 \right] \quad (8.15)$$

Integrating again, from eqn (8.14),

$$A_{FB} = \frac{3}{4} \frac{2c_V^{(e)} c_A^{(e)}}{\left(c_V^{(e)} \right)^2 + \left(c_A^{(e)} \right)^2} \frac{2c_V^{(f)} c_A^{(f)}}{\left(c_V^{(f)} \right)^2 + \left(c_A^{(f)} \right)^2} = \frac{3}{4} A_e A_f \quad (8.16)$$

where

$$A_e = \frac{2c_V^{(e)} c_A^{(e)}}{\left(c_V^{(e)} \right)^2 + \left(c_A^{(e)} \right)^2}, \quad A_f = \frac{2c_V^{(f)} c_A^{(f)}}{\left(c_V^{(f)} \right)^2 + \left(c_A^{(f)} \right)^2}$$

What electroweak information can be obtained from these measurements? σ and A_{FB} for a particular final state f give two equations involving $c_V^{(f)}$ and $c_A^{(f)}$, and, in principle, these can be solved to give both $c_V^{(f)}$ and $c_A^{(f)}$ separately.¹⁹ On a plot of c_V versus c_A , $\sigma \propto c_V^2 + c_A^2$ is a circle and $\sigma A_{FB} \propto c_V c_A$ has a hyperbolic dependence. Taken together, these should enable the extraction of both c_V and c_A for the fermion f (see Exercise 8.6). This can be done for each final state f . In addition, note that σ and A_{FB} also depend on the electron values for $c_V^{(e)}$ and $c_A^{(e)}$ to be completely unravelled. These must either be taken from other experiments or obtained using separate measurements from $\tau^+\tau^-$ events.

¹⁸The angle θ is the angle between the incoming electron direction and the outgoing lepton or quark, or between the incoming positron and outgoing antilepton or antiquark.

¹⁹Measuring σ correctly requires accurate knowledge of the luminosity, which is described in Section 8.4.10.

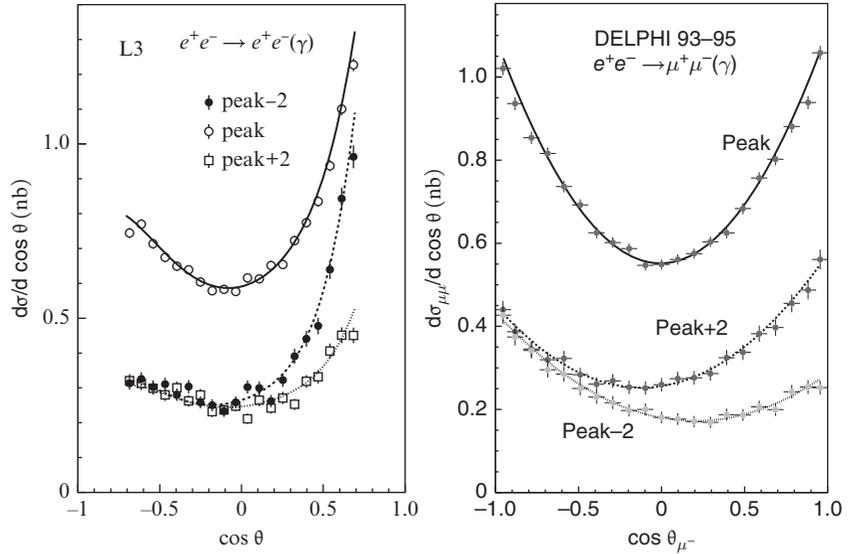


Fig. 8.12 Cross section as a function of θ , the angle between the incoming electron and outgoing lepton. The two plots are the results from two different LEP detectors; the three curves show the three main beam energies at and near the Z pole, where LEP was run. From [17].

We have derived the A_{FB} formula assuming all the interactions are mediated by the Z , but we also need to account for the γ exchange diagram and the interference between the two diagrams. Figure 8.12 shows the differential cross section as a function of $\cos \theta$ when LEP was run on the Z pole and 2 GeV either side. The variation of A_{FB} as a function of beam energy is shown in Fig. 8.13.

In summary, then, from the measured quantities σ_{ee} , $\sigma_{\mu\mu}$, $\sigma_{\tau\tau}$, A_{FB}^e , A_{FB}^μ , and A_{FB}^τ , and some information from τ decay, we obtain $c_V^{(e)}$, $c_V^{(\mu)}$, $c_V^{(\tau)}$, $c_A^{(e)}$, $c_A^{(\mu)}$, and $c_A^{(\tau)}$. From the data, we see the following:

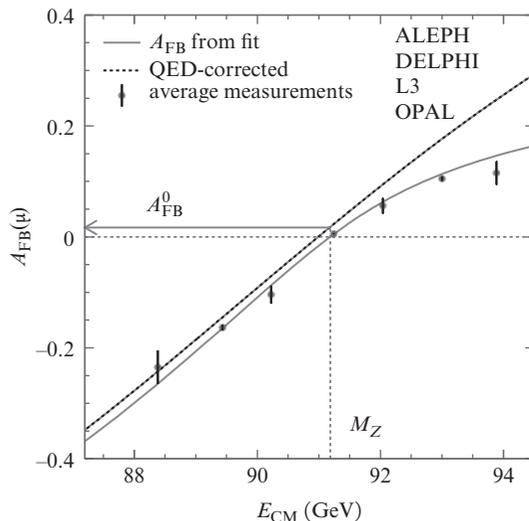


Fig. 8.13 A_{FB} as a function of LEP beam energy [17].

- (1) Universality: $c_V^{(e)} = c_V^{(\mu)} = c_V^{(\tau)}$ and also $c_A^{(e)} = c_A^{(\mu)} = c_A^{(\tau)}$ to high precision. All three types of lepton couple to the Z^0 with the same strength.
- (2) $\sin^2 \theta_W = 0.23189 \pm 0.00024$ (this is the combined value from all LEP techniques).

8.4.10 LEP luminosity measurement

Accurate data on the LEP luminosity is essential for cross-section measurements. Recall that the integrated luminosity L is defined by the expression $N_i = L\sigma_i$ where N_i is the number of events from a given process i that occur in the detector (once detector effects like trigger efficiency and acceptance have been corrected for) and σ_i is the cross section. L is the same for all processes—it depends on the features of the accelerator and how well it is working (e.g. how well the two beams are steered into each other at a particular interaction region). Luminosity is measured using a process j with a high rate and for which we know how to calculate the cross section. For LEP, the channel j used to measure the luminosity was low-angle electron–positron scattering (Bhabha scattering), for which the QED single-photon-exchange diagram dominates. There are only very small contributions to this from diagrams with Z^0 or involving annihilation, and the theoretical uncertainty is below 0.1%. Bhabha scattering events are measured with small calorimeters situated very close to the beam line on each side of the detectors, several metres each side of the interaction point. The events are counted when two showers with the appropriate energy are seen within a radius region about the beam axis of typically $6\text{ cm} < R < 15\text{ cm}$. This cut is made on one side only to reduce sensitivity to the movement of the interaction point. Using the numbers of events N_i and N_j and the formulae $L = N_j/\sigma_j$ and $\sigma_i = N_i/L$, we obtain the cross section in which we are interested.²⁰

²⁰In practice, luminosity measurements will be recorded as a function of date and time after correction for acceptance and radiative corrections (higher-order QED processes) and made available for all analyses.

8.4.11 Measurements at LEP2, above $\sqrt{s} = M_Z$

Superconducting cavities were added to the LEP ring from 1995 onwards and the accelerator was run at increasing energies as more cavities were added (with increasing beam energy, the energy lost by synchrotron radiation each turn becomes greater and more cavities are needed to replace it). The cross section as a function of beam energy is shown in Fig. 8.14. The maximum energy reached was a little over $\sqrt{s} = 200\text{ GeV}$ and the last run was in 2000. The main studies at these higher energies were searches for new particles: Higgs and SUSY being the most popular. Precision measurements of the triple gauge coupling were also made.²¹ The higher energy also facilitated other detailed studies of the W , which complemented those done at $\bar{p}p$ colliders.

A useful class of events at LEP2 is ‘radiative return to the Z^0 ’ or initial-state radiation (ISR). This happens when either the beam electron or positron emits a bremsstrahlung photon (Fig. 8.15) and loses enough

²¹These studies required the CMS energy to be above $2M_W$.

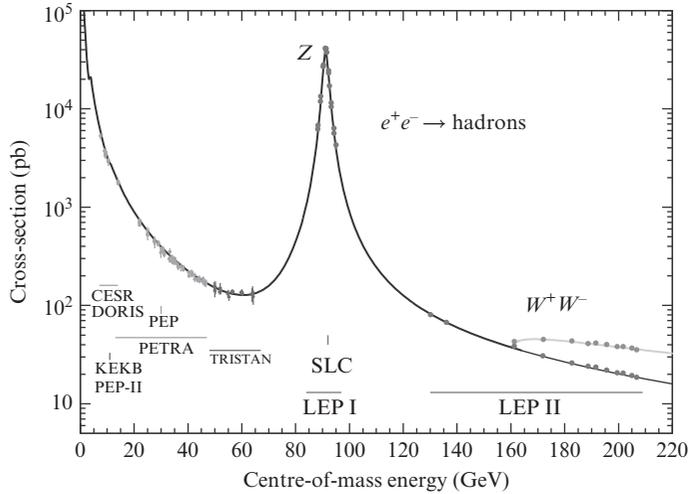


Fig. 8.14 Cross section versus \sqrt{s} near and above the Z pole. From [17].

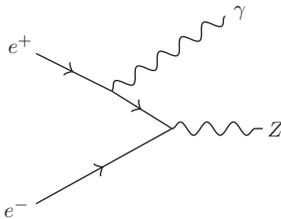


Fig. 8.15 One Feynman diagram for initial-state radiation (ISR).

energy for the subsequent collision to have the right energy to make an on-mass-shell Z^0 . This is very significant, because the Z^0 resonance is so large. Measuring cross sections and A_{FB} above the Z peak provides more consistency checks of the electroweak model.

8.4.12 $W^+ W^-$ production

This is a good example for studying some of the important aspects of a physics analysis either at LEP or at a hadron collider. The two main aspects are (1) choice of final states and (2) combinatorics. $W^+ W^-$ events can come in three different types depending on whether each of the W s decays to a pair of quarks or to a charged lepton and a neutrino. For events with quarks in the final state, we need to run a ‘jet’ algorithm to assign measured charged particle tracks and calorimeter energies not associated with tracks to a particular jet.²²

²²Jet-finding algorithms are discussed in Section 13.3. A nice example of a two-jet event is shown in Fig. 8.6(d).

- **4 jets:** We get the 3-momentum components of each of 4 jets and we know the beam energy (a total of 13 numbers), so we can do a constrained fit of the event. The things we do not know are the two decay angles of the Z and of each of the two W s. We also leave the masses of the two W s free in the fit. It is therefore a 5-constraint fit (13 numbers, 8 unknowns). There is also a combinatorial problem—we do not know which pairs of jets go together—so we try each of the three combinations in turn: 12, 34; 13, 24; 14, 23. For all three combinations, we make a scatter plot of the mass of one W versus the mass of the other and hope to see a peak where the real W is. Intuition tells us that the false combinations are likely to be scattered widely about the plot, not forming a mass peak, and a Monte Carlo simulation can be used to provide more quantitative information about the distribution of the false combinations.

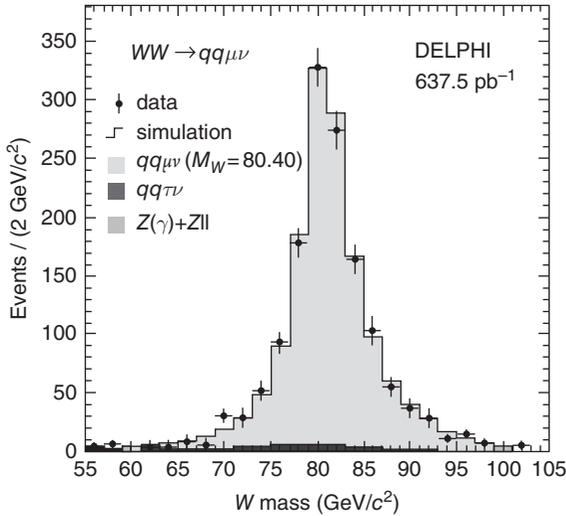


Fig. 8.16 W mass distribution from the 2-constraint fit for $W \rightarrow \mu\nu_\mu$, $W \rightarrow q\bar{q}$. Similar distributions are obtained for the $qqq\bar{q}$, $e\nu qq$, and $\tau\nu qq$ channels. From [5].

- **2 jets, l, ν :** There are fewer constraints here, because the neutrino is undetected and deprives us of 3 of the 13 numbers we had in the 4-jet case. Nevertheless, this is a 2-constraint fit and is therefore fairly powerful. There is no combinatorial problem here—the detector tells us which is the lepton. This analysis is only done when the lepton is an electron or a muon. If the lepton is a τ , the τ decay must involve at least one other neutrino, a τ neutrino. There is still sufficient information for a 1-constraint fit.

An example of a W mass fit from the $\mu\nu_\mu q\bar{q}$ channel [5] is shown in Fig. 8.16. This technique is a good way to measure the mass of the W . Each individual event gives two independent measurements of the mass and all we need do is take an average to get the mass. If we want to be more sophisticated, we can estimate the error on each mass measurement (from the errors on the track and calorimeter measurements) and do a weighted average. The spread of the individual M_W measurements gives us the opportunity to measure the width of the W (the spread is determined by the natural width and the experimental resolution). The result is $\Gamma_W = 2.48 \pm 0.41$ GeV. The prediction from electroweak theory is 2.077 GeV, which is consistent. This is not as precise, however, as the measurements from the Tevatron (the CDF and D0 experiments) discussed in Section 8.5.3.

8.4.13 $\sigma(e^+e^- \rightarrow W^+W^-)$

The three diagrams contributing to the $e^+e^- \rightarrow W^+W^-$ cross section are shown in Fig. 8.17. All three are needed to give the cancellation required to avoid a divergent theoretical result for this cross section as the energy increases. It is therefore very satisfying to see that the measured values agree with the theory when all three diagrams are present

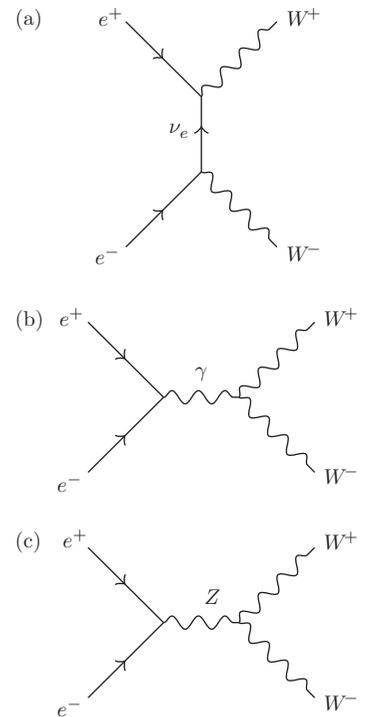


Fig. 8.17 Three lowest-order Feynman diagrams for W -pair production: (a) t -channel ν_e exchange; (b) s -channel γ exchange; (c) s -channel Z exchange.

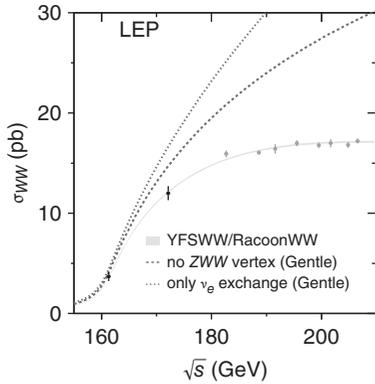


Fig. 8.18 The \sqrt{s} energy dependence of $\sigma(Z \rightarrow WW)$ near threshold, from which the W mass may be determined. The lower curve, which follows the data, shows the predicted cross section including all three of the diagrams shown in Fig. 8.17, and the other curves show the prediction if some of these diagrams are omitted. From [16].

²³The W was discovered before the Z because the product of cross section and branching ratio is an order of magnitude bigger than that for the Z .

²⁴The fact that the W and Z signatures were so clean was a major turning point in the subject. Previous prejudice was that hadron colliders were too ‘dirty’. This change of attitude opened the way to the construction of the LHC.

but completely disagree with a calculation neglecting the triple-gauge-boson coupling. This provides conclusive evidence for the existence of the triple-gauge-boson couplings γW^+W^- and ZWW , with the rates agreeing with those predicted by the unified electroweak theory.

Also, the threshold energy for W^+W^- pair production gives a separate measurement of the mass of the W . The shape of $\sigma(Z \rightarrow W^+W^-)$ as a function of \sqrt{s} just above threshold is shown in Fig. 8.18. The result for the mass of the W from LEP (combining the cross-section threshold measurement and the individual event reconstruction method described above) is $M_W = 80.39 \pm 0.09$ GeV. The Tevatron experiments provide a better measurement, as will be discussed in Section 8.5.2.

8.5 W and Z physics at hadron colliders

8.5.1 W and Z discovery

The W and Z were discovered in 1983 at the CERN $\bar{p}p$ collider. The accelerator physics required to produce sufficiently intense \bar{p} beams is reviewed in Chapter 3. The largest branching ratios for W and Z are to hadronic final states, but these are very difficult to study in a hadron collider because of the very large cross section for QCD jets (see Chapter 9). Therefore, the experiments focused on the relatively clean leptonic decay modes: $W \rightarrow e\nu_e$, $W \rightarrow \mu\nu_\mu$, and $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$.²³ The key detector feature that enabled the W discovery was the use of ‘hermetic’ detectors, which allowed the neutrino transverse momentum to be determined from the measured missing transverse energy (E_T^{miss} ; see Chapter 4). As the mass of the W was predicted to be very large (~ 80 GeV), the signature was a high-transverse-momentum lepton (electron or muon) and a high value for E_T^{miss} . While there are backgrounds from QCD that can produce ‘fake’ electrons or muons in the detector, they would not usually have such a large energy, nor would a QCD event produce such a large value of E_T^{miss} . The signature for a Z is an e^+e^- or $\mu^+\mu^-$ pair with invariant mass consistent with that of the Z , resulting in a narrow resonance on top of a low background. The measured W and Z masses were in very good agreement with the predictions from electroweak theory.²⁴ Much higher-precision measurements of the Z mass were made at LEP, as discussed in Section 8.4.

Test of V–A

We will need to assume that the W decays are via a V–A coupling in order to be able to use their distinctive signature quantitatively. The most direct and in some sense simplest test of the V–A theory of weak decays is provided by measuring the angular distribution of the charged leptons resulting from W decay. We can calculate the CMS angular distribution, as will be discussed shortly. As the W s are produced with finite longitudinal momentum, we need to boost the measured event to

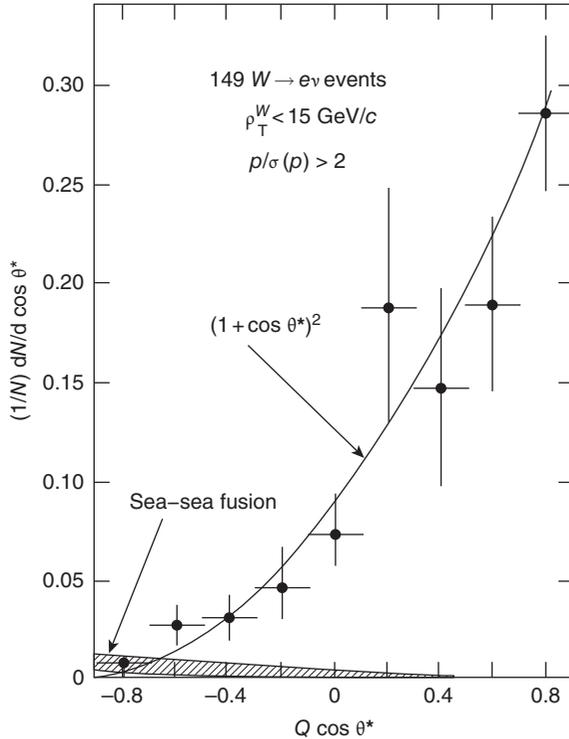


Fig. 8.19 Measured angular distribution in $W \rightarrow e\nu_e$ decays [15].

the $q\bar{q}$ CMS. We do not directly determine the longitudinal momentum of the neutrino, but it can be determined up to a quadratic ambiguity by using the W mass constraint. The measured distribution [15] is shown in Fig. 8.19.

8.5.2 W mass determination at the Tevatron

The current best measurement of the W mass comes from the Tevatron $\bar{p}p$ collider, which had a CMS energy of 1.96 TeV. The mass of the Z was measured very precisely at LEP. Combining precision measurements of the W and Z bosons is interesting because the result provides a window on possible higher-mass particles via radiative corrections (see Section 7.4.3).

The decay modes $W \rightarrow l\nu$ with $l = e$ or $l = \mu$ are used for the W mass determination, since they have very clean signals. The neutrino is not directly detected; however, the transverse component of the neutrino momentum can be determined from the missing transverse momentum in the event. Too much energy is lost in the beam pipes for the longitudinal component to be determined. Also, the quarks carry unknown fractions x and \bar{x} of the proton and antiproton momenta, and hence the invariant mass of the W cannot be computed from the final-state particles in an individual event. However, the mass of the W can be determined

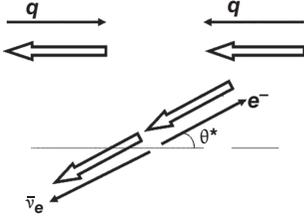


Fig. 8.20 Spin structure in $q\bar{q} \rightarrow W^- \rightarrow e^- \bar{\nu}_e$ events.

from the data on a statistical basis. The W bosons are produced via the parity-violating $V-A$ interaction. We will assume that the quarks (antiquarks) come from the p (\bar{p}) because the typical values of the proton momentum fraction x (antiproton momentum fraction \bar{x}) are quite large. All the quarks and leptons are ultrarelativistic, so the $V-A$ interaction, which couples left-handed particles (right-handed antiparticles), will result in negative-helicity particles (positive-helicity antiparticles). The spin structure in W production and decay is shown in Fig. 8.20.

The W^- has a spin along the proton beam direction of $J_z = -1$ and the $l^- \bar{\nu}_l$ system has $J_{z'} = -1$ along an axis pointing in the direction of the l^- . The angular distribution in the W CMS is then given simply by the rotation matrix (see Chapter 2):

$$\frac{dN}{d \cos \theta^*} = N_0 [d_{-1,-1}^1(\cos \theta^*)]^2 = \frac{N_0}{4} (1 + \cos \theta^*)^2 \quad (8.17)$$

where N_0 is a normalization constant. From the chain rule, we can change variable to the component of the momentum of the electron perpendicular to the beam, p_{T}^e :

$$\frac{dN}{dp_{\text{T}}^e} = \frac{dN}{d \cos \theta^*} \frac{d \cos \theta^*}{dp_{\text{T}}^e} \quad (8.18)$$

with

$$\cos \theta^* = \sqrt{1 - (\sin \theta^*)^2} = \sqrt{1 - \left(\frac{2p_{\text{T}}^e}{M_W}\right)^2} \quad (8.19)$$

Differentiating gives

$$\frac{d \cos \theta^*}{dp_{\text{T}}^e} = - \frac{4p_{\text{T}}^e/M_W^2}{\sqrt{1 - \left(\frac{2p_{\text{T}}^e}{M_W}\right)^2}} \quad (8.20)$$

²⁵The peak arises from the Jacobian of the change of variables and is often called the ‘Jacobian peak’.

Substituting from eqn 8.20 into eqn 8.18 gives²⁵

$$\left| \frac{dN}{dp_{\text{T}}^e} \right| = N_0 \frac{p_{\text{T}}^e}{M_W^2} \frac{\left[1 + \sqrt{1 - \left(\frac{2p_{\text{T}}^e}{M_W}\right)^2} \right]^2}{\sqrt{1 - \left(\frac{2p_{\text{T}}^e}{M_W}\right)^2}} \quad (8.21)$$

²⁶At collider energies, the masses of the leptons (e and μ) are negligible compared with their momenta, and therefore we can use the massless approximation $p = E$.

From this expression, it can be seen that the p_{T} spectrum will be peaked towards its upper endpoint at $M_W/2$. In principle, the W mass can be determined by fitting the measured charged-lepton p_{T} distribution to eqn 8.21.²⁶ However, the shape of the distribution is distorted by the distribution of the transverse momentum of the W , P_{T}^W . Therefore, it

is convenient to define the transverse mass M_T in a similar way to the invariant mass but only considering transverse components:

$$M_T^2 = (p_{l,T} + p_{\nu,T})^2 - |\mathbf{p}_{l,T} + \mathbf{p}_{\nu,T}|^2 \quad (8.22)$$

This can also be expressed as (see Exercise 8.5)

$$M_T^2 = 2p_T^l p_T^\nu (1 - \cos \Delta\phi) \quad (8.23)$$

where $\Delta\phi$ is the angle between \mathbf{p}_T^l and \mathbf{p}_T^ν , the transverse momenta of the lepton and neutrino, respectively.

Experimental aspects of measuring M_W from the M_T distribution

For W bosons produced with no transverse momentum, $M_T = 2p_T^l$ and to lowest order the effect of non-zero values of P_T^W will not change the value of M_T ; hence there should be an endpoint at M_W . The distribution is smeared by the finite width of the W and by experimental resolution. However, if these effects are taken into account, the data can be used in a fit to determine M_W . There are also several systematic uncertainties that must be understood before a precision measurement of M_W can be made.

One of the most important of these is the energy scale for electrons and muons; this can be constrained by fitting the l^+l^- invariant mass (where l is an e or a μ) to the peaks from Z , Υ , and J/ψ decays. Using the known masses of these resonances, the energy scale can be calibrated, allowing for any nonlinearity. As $M_Z > M_W$, it is essential to fix the nonlinearity as well as the overall energy scale of the detector.

Another key systematic uncertainty is the measurement of p_T^ν by the method of missing transverse momentum. This can be constrained by studying the apparent missing transverse energy in $Z \rightarrow l^+l^-$ decays and looking at the agreement between the values of p_T^Z inferred from the accurately measured l^+l^- system and the hadronic recoil. The axes perpendicular and parallel to the bisector of the l^+l^- system are defined as shown in Fig. 8.21. For a perfect detector, the transverse momentum of the l^+l^- system would be balanced by the hadronic recoil u . The l^+l^- system is measured with far better precision than that of the hadronic recoil. Therefore, the distribution of $p_T^l + u$ can be used to determine the hadronic response. An example of such a plot from the CDF experiment [1] is shown in Fig. 8.22, which shows the mean value of $p_\eta^l + u_\eta$ as a function of p_T^l . The quantity is projected onto the η axis because the experimental error in the value of p_η^l is mainly due to errors in the measurements of angles, rather than energy. Measurements of angles from the tracking detector are very precise, so the measurement errors in this quantity are greatly reduced. The value of this quantity would be 0 for an ideal detector and the fact that the mean value is positive is due to energy loss outside the acceptance of the detector and in cracks between calorimeter cells.

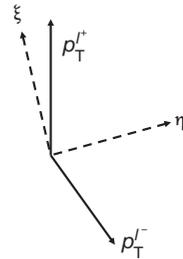


Fig. 8.21 Definition of the axes parallel and perpendicular to the bisector of the l^+l^- system.

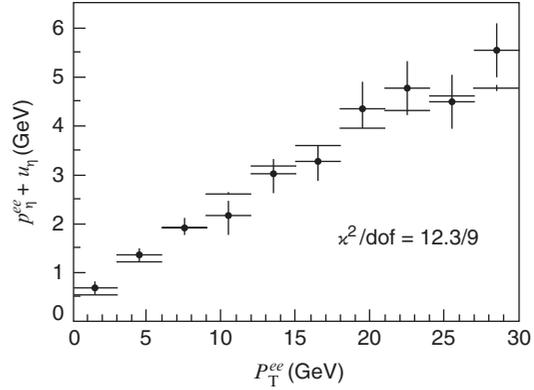


Fig. 8.22 Calibration of the hadronic response by measuring the mean value of $p_{\eta}^{ll} + u_{\eta}$ versus p_T^{ll} as measured by CDF in $Z \rightarrow e^+e^-$ decays [1].

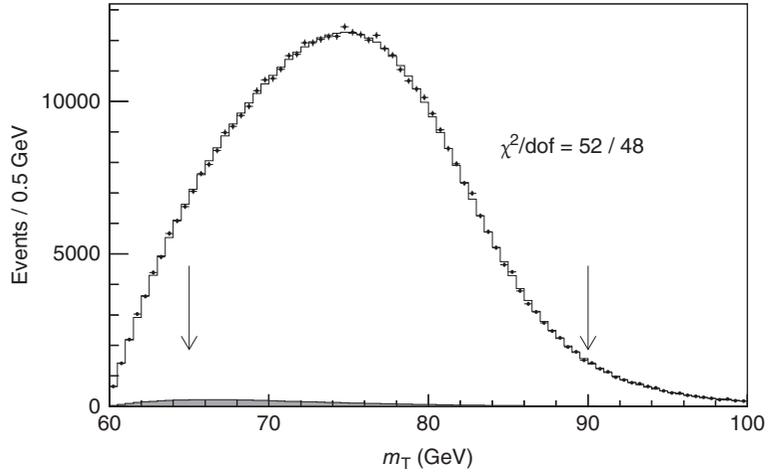


Fig. 8.23 W mass fit at CDF in the $W \rightarrow e\nu_e$ channel [3].

The M_T distribution for $W \rightarrow \mu\nu_{\mu}$ shows a very clean Jacobian peak (see Fig. 8.23). From the fits to the M_T spectra from $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_{\mu}$, the W mass is determined to be $80\,387 \pm 19$ MeV, which is the most precise measurement to date [3]. The current (2014) world average value [115] is $80\,385 \pm 15$ MeV.

8.5.3 Width of the W

The total width of the W was measured at CDF and D0 using two methods to compare it with the value predicted by the electroweak theory of $\Gamma_W = 2.077$ GeV. The direct technique is to extract it from the same transverse mass distribution as described above. The distribution will be modified depending on the W width. The value obtained is $\Gamma_W = 2.11 \pm 0.32$ GeV.

The indirect method of obtaining the total W width involves using some information from LEP. The width is obtained by measuring the production cross-section ratio²⁷

$$\frac{\sigma(\bar{p}p \rightarrow W \rightarrow l\nu)}{\sigma(\bar{p}p \rightarrow Z \rightarrow \bar{l}l)} = \frac{\sigma(\bar{p}p \rightarrow W)}{\sigma(\bar{p}p \rightarrow Z)} \times \frac{\Gamma(W \rightarrow l\nu)}{\text{BR}(Z \rightarrow \bar{l}l)} \times \frac{1}{\Gamma_W} \quad (8.24)$$

The other numbers in the formula are obtained as follows: the ratio $\sigma(\bar{p}p \rightarrow W)/\sigma(\bar{p}p \rightarrow Z)$ is predicted by modelling;²⁸ $\Gamma(W \rightarrow l\nu)$ is predicted from the electroweak theory; $\text{BR}(Z \rightarrow \bar{l}l)$ was measured at LEP.

The result obtained is $\Gamma_W = 2.062 \pm 0.059$ GeV. The importance of this measurement is that the total W width is sensitive to any particles into which the W might decay, including some that we might not otherwise have discovered yet. This indirect method produces a much more precise check than either the direct method or the measurements from LEP. Everything is consistent with the prediction from the electroweak theory.

²⁷Measuring a ratio of cross sections is easier than measuring a correctly normalized single cross section, since many systematic effects will cancel out.

²⁸Again a ratio is easier to predict.

8.6 Top-quark physics

The top quark was expected ever since the b quark was discovered in 1977. From the measurements at LEP of $Z \rightarrow b\bar{b}$, the b quark was confirmed to be a member of a weak isodoublet. While these results gave strong indications that the top quark must exist, it was the combination of the precise electroweak data with the radiative corrections that gave a prediction for the top-quark mass of ~ 170 GeV (see Chapter 7). We will review the discovery of the top quark in Section 8.6.1 and then consider how to make precision measurements of its mass in Section 8.6.2.

8.6.1 Top-quark discovery

Given the high mass, the only accelerator at the time capable of producing the top quark was the Tevatron. The largest cross sections for producing top quarks at the Tevatron are pair production from either gg or $q\bar{q}$ initial states (see Figure 8.24)

The top quark decays very rapidly, with a branching ratio of essentially 100% into a W and a b quark. Since the mass of the top is much larger than that of the W , the W from top decay is on mass shell and it decays through all its decay modes with the known branching ratios. Therefore, it is easy to work out the fraction of events with different final-state topologies. All events will have two b -jets. The decays of the two W s leading to final states that can be identified above the backgrounds are listed in Table 8.2.

The main Standard Model background is W + jets from higher-order QCD corrections to W production. In general, this type of background event will not contain b quarks. Therefore, identifying jets containing b quarks is a critical technique for separating the top signal from

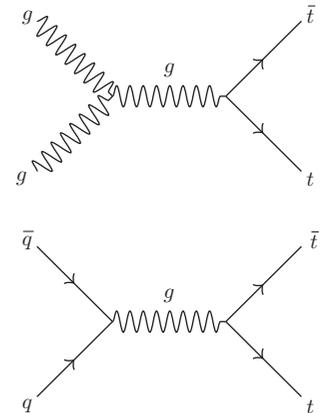


Fig. 8.24 Lowest-order Feynman diagrams for $t\bar{t}$ production.

Final state	Fraction
$e\mu$	1/81
e^+e^- or $\mu^+\mu^-$	2/81
e -jets or μ -jets	12/81
All jets	36/81

Table 8.2 Some $t\bar{t}$ final-state topologies.

backgrounds. There are two strategies for identifying b quarks (called ‘ b -tagging’):

- Use semileptonic decays of the b quark to search for leptons (in practice only e or μ). Because of the relatively high mass of the b quark, these leptons will tend to have high transverse momentum with respect to the jet axis.
- Use the relatively long lifetime of the b quark. The b quark fragments into B mesons or baryons before it decays. The lifetime is of the order of a picosecond, so, allowing for time dilation, the decay vertices of B hadrons can be displaced by distances of the order of millimetres.

The first technique is limited by the semileptonic branching ratio, so lifetime tagging is more powerful. This requires very high-precision tracking close to the interaction point. As B hadrons will typically decay inside the beam pipe, very high-precision detectors of very low mass to minimize multiple scattering are required. The only practical technique that can satisfy both requirements is based on the use of silicon detectors (see Chapter 4). Starting from the interaction point, the beam pipe within the vertex detector is made from beryllium ($Z = 4$), since this has a very long radiation length (which minimizes distortion of charged-particle tracks due to multiple scattering). The active detector components consist of layers of silicon strip detectors around the beam pipe. The first layer is mounted on the beam pipe to minimize the distance to the beam line. This reduces extrapolation errors and the effect of multiple scattering in the beam pipe. The transverse impact parameter is defined as the closest approach of an extrapolated track to the primary vertex location in the plane transverse to the beam line. Short-lived hadrons from light-quark jets produce a Gaussian distribution and the long-lived B hadrons generate an exponential tail at high values of the impact parameter.

The power of b -tagging to identify the top-quark signal is illustrated in Fig. 8.25 [7], which shows the number of events with W 's and various numbers of high-transverse-momentum jets before and after b -tagging. The signal events from $t\bar{t}$ events should ideally contain 4 jets, but this can be distorted by instrumental effects. The $W + \text{jet}$ events will decrease rapidly with increasing number of jets, because each additional jet has a penalty of the order of α_s . A very clear signal is visible for 4 jets after b -tagging, as shown in Fig. 8.25.

8.6.2 Top-quark mass measurement

Once a clean signal for $t\bar{t}$ has been established, these events can be used to fit the top-quark mass m_t by reconstructing the decay products of the t and \bar{t} . This procedure is difficult, since there is no way of uniquely identifying which jet came from the t and which from the \bar{t} . Many possible combinations have to be tried in turn and some algorithm used to select the most likely combination. In the ‘all-jets’ channel, all the decay

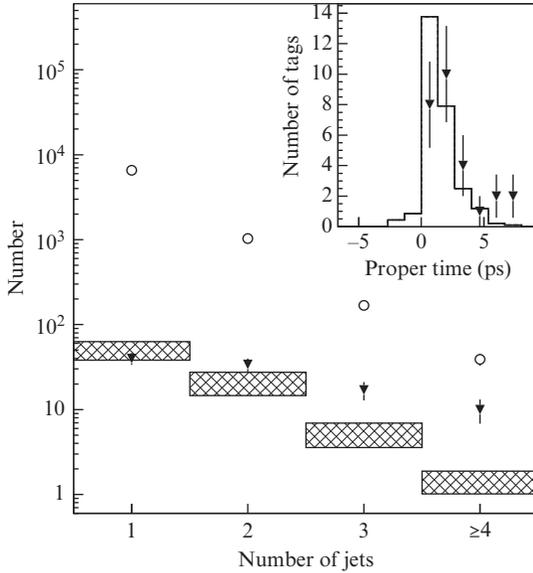


Fig. 8.25 Distribution of the number of jets before and after b -tagging as measured by CDF. The circles (triangles) are the data before (after) b -tagging and the shaded boxes represent the background estimates after b -tagging. From [7].

products of the t and \bar{t} are measured directly, but this channel suffers from a large background.

The semileptonic channels are cleaner, but in a hadron collider only the momentum of the neutrino transverse to the beam direction can be reconstructed. The unknown longitudinal momentum of the neutrino can be determined by exploiting the constraint from the decay $W \rightarrow l\nu$:

$$(M_W)^2 = (E_\nu + E_l)^2 - (p_\nu + p_l)^2 \quad (8.25)$$

The neutrino momentum is split into a transverse component $p_{\nu,T}$, which can be inferred from the measured missing transverse energy, and an unknown longitudinal component $p_{\nu,L}$. This gives

$$(M_W)^2 = \left(\sqrt{p_{\nu,T}^2 + p_{\nu,L}^2} + E_l \right)^2 - (p_{\nu,T} + p_{l,T})^2 - (p_{\nu,L} + p_{l,L})^2 \quad (8.26)$$

Using the known mass of the W , the value of $p_{\nu,L}$ can be determined up to a twofold ambiguity by solving the quadratic equation 8.26. An example of a fit to the top mass is shown in Fig. 8.26 [2].

The combined result of the top-quark mass measurements from the Tevatron CDF and D0 experiments is $m_t = 173.20 \pm 0.51(\text{statistical}) \pm 0.71(\text{systematic}) \text{ GeV}$ [131]. The top-quark measurements at LHC will be reviewed in Chapter 13.

8.6.3 Top-quark production cross sections

As well as measuring the mass of the top quark, another key measurement is the top-quark production cross section. The Tevatron collider measurements [115] for the total top production cross section

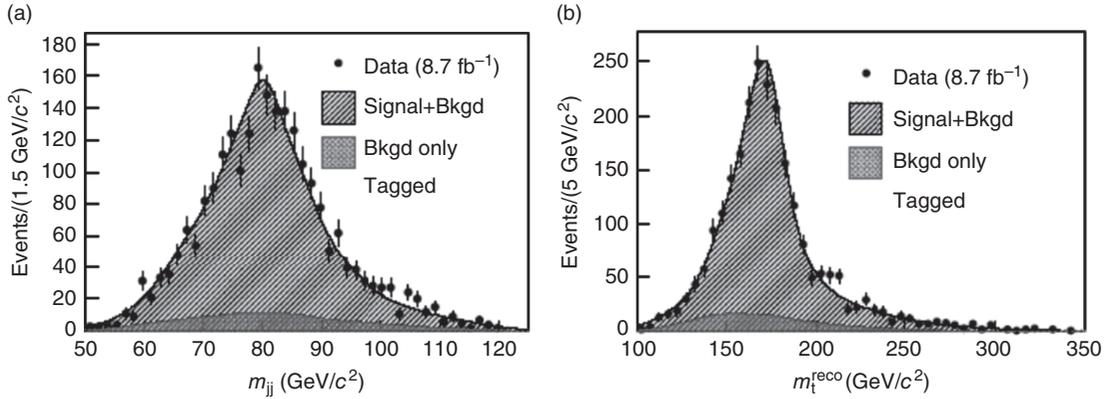


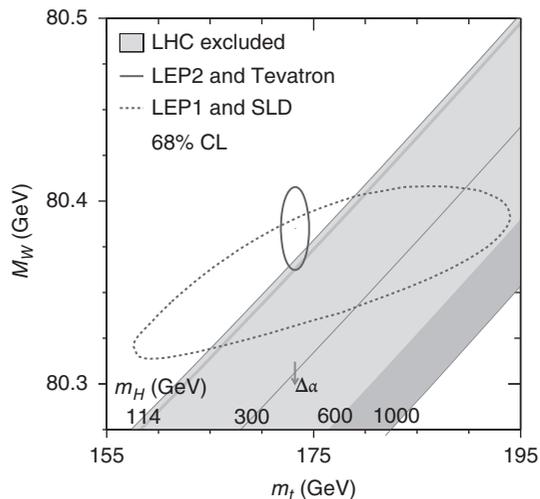
Fig. 8.26 Top-quark mass fits from the CDF experiment. (a) Reconstructed dijet mass, which shows the expected peak at M_W . (b) Distribution of reconstructed top-quark masses. From [2].

$\sigma_{\text{tot}}(p\bar{p} \rightarrow t\bar{t})$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ are $\sigma_{t\bar{t}}^{\text{D}0} = 7.56^{+0.63}_{-0.56}$ (statistical + systematic) pb and $\sigma_{t\bar{t}}^{\text{CDF}} = 7.50 \pm 0.48$ (statistical + systematic) pb. The methodology for calculating cross sections for Standard Model processes will be described in Chapter 9.

8.7 Summary

All electroweak data are consistent with the Standard Model at a precision sufficient to be sensitive to radiative corrections. This enables us to place limits on masses and couplings of new particles. For example, using the measured masses of the top quark and the W boson within the context of the Standard Model, we can predict a value for the mass of the Higgs boson. The results are shown in Fig. 8.27. Combining this

Fig. 8.27 The mass of the W versus the mass of the top quark, with results from direct measurements and indirect determinations from Standard Model fits. The region below $m_H < 114 \text{ GeV}$ is excluded by direct searches at LEP. The small ellipse shows the Standard Model fit to all the precision data, including the direct measurements of the W and top-quark masses, whereas the larger dashed ellipse includes only the precision data collected near the Z pole. The shaded diagonal bands show the Standard Model expectations for different ranges of the Higgs boson mass. From [97].



analysis with the results of the direct Higgs search at LEP (see Chapter 12), the 95% confidence level allowed range [97] for the Higgs mass is $114 \text{ GeV} < m_H < 149 \text{ GeV}$. The discovery of a Higgs boson in this mass range is discussed in Chapter 12.

Chapter summary

- Experiments confirmed the V–A structure for the charged-current weak interactions.
- Neutral currents were detected, which was the first success for the unified electroweak theory.
- The W and Z bosons were discovered at the expected masses.
- The electroweak theory was tested with very high precision at LEP running on the Z boson.
- Many precision measurements allowed the Standard Model to be probed at the level of the radiative corrections, which allowed the successful prediction of the top-quark mass.
- The measured top-quark mass combined with the other precision data allowed narrowing of the allowed range for a Standard Model Higgs

Further reading

- Cahn, R. N. and G. Goldhaber, G. (2009). *The Experimental Foundations of Particle Physics* (2nd edn). Cambridge University Press. Chapters 6 and 12–15 give a summary and reprints of key papers on weak interactions and electroweak unification.
- Particle Data Group (2014). Review of Particle Physics. *Chin. Phys. C*, **38**, 090001. In the section ‘Standard Model and related topics’, the review article ‘Electroweak model and constraints on new physics’ gives a full discussion of the Standard Model fits to all the precision electroweak data.
- Behnke, O. et al. (Eds.) (2013). *Data Analysis in High Energy Physics*. Wiley. A good book on advanced data analysis techniques.

Exercises

- (8.1) Consider the reaction used in the determination of the neutrino helicity: $^{152}\text{Eu}(J = 0) + e^- \rightarrow ^{152}\text{Sm}^*(J = 1) + \nu_e$, followed by $^{152}\text{Sm}^*(J = 1) \rightarrow ^{152}\text{Sm}(J = 0) + \gamma$. The energy released in the electron capture of the ^{152}Eu is 840 keV.
- (a) Show that the γ emitted in $^{152}\text{Sm}^*$ decays in which the nucleus is at rest have too low energy to undergo the inverse reaction $\gamma + ^{152}\text{Sm}(J = 0) \rightarrow ^{152}\text{Sm}^*(J = 1)$. The width of the resonance at 960 keV in $^{152}\text{Sm}^*$ is about 3 eV.

- (b) Show that if the γ is emitted in the forward direction with respect to the recoiling Sm^* , then it will have sufficient energy to undergo $\gamma + {}^{152}\text{Sm}(J=0) \rightarrow {}^{152}\text{Sm}^*(J=1)$.
- (c) Consider the 4 combinations of the spins of the electron and the neutrino to show that for the allowed cases, the helicity of the photon is equal to that of the neutrino.

Hint: The spin of the photon projected on its direction of propagation can only be +1 or -1.

- (8.2) Justify eqn 8.1.

Hint: first consider the allowed couplings of L and R leptons and use the Standard Model relation for the L and R coupling constants in terms of the weak mixing angle (see eqn 7.40).

- (8.3) Justify the condition for the electron angles and energies given by eqn 8.2.
- (8.4) If the neutrinos cannot be directly detected, how can one detect $Z \rightarrow \nu\bar{\nu}$ events in e^+e^- ? Draw an appropriate Feynman diagram and discuss the resulting detector requirements.

Hint: Consider the effects of initial-state radiation (ISR).

- (8.5) Show that eqn 8.23 can be derived from eqn 8.22.
- (8.6) Calculate the forward-backward asymmetry A_{FB}^{ff} and the total cross-section for the reactions
- $e^+e^- \rightarrow \mu^+\mu^-$
 - $e^+e^- \rightarrow b\bar{b}$

- (8.7) Determine the angular distribution given in eqn 8.14.

- (8.8) The cross section for inverse beta decay is given by

$$\sigma(\bar{\nu}_e + p \rightarrow e^+ + n) = \frac{4G_{\text{F}}^2}{\pi} p^2$$

Derive this cross section from Fermi's Golden Rule, ignoring spin and assuming that the matrix element $M_{\text{if}} = 2G_{\text{F}}$. In a fission reactor, the energy released per fission is ~ 200 MeV and there are 6 antineutrinos released. The average energy of the anti-neutrinos is approximately 2 MeV. What is the minimal power of a nuclear fission reactor to have three interactions per hour in a 200 kg water tank at 10 m distance

- (8.9) At a given neutrino energy, the ratio of antineutrino to neutrino cross-sections for scattering on electrons is

$$\sigma(\bar{\nu}_\mu)/\sigma(\nu_\mu) = 0.85$$

Deduce a value for $\sin^2 \theta_{\text{W}}$.

- (8.10) In an experimental run at LEP at the Z^0 peak energy, the integrated luminosity was 23.955 pb^{-1} . After corrections for background and detector efficiencies, a total of 993 797 hadronic events and 47 838 $\mu^+\mu^-$ events were obtained. Stating your assumptions, calculate the number of neutrino flavours and comment on your result.

- (8.11) The cross section for $e^+e^- \rightarrow W^+W^-$ just above threshold is given by

$$\sigma_{\text{T}} \simeq \frac{2G_{\text{F}}^2 M_{\text{W}}^4}{\pi s} \sqrt{1 - \frac{4M_{\text{W}}^2}{s}}$$

where G_{F} is the Fermi coupling constant, M_{W} the W mass, and s the CMS energy squared, all in natural units. Express σ_{T} in terms of s and β . With e^+ and e^- beams of energy 80.65 GeV, a cross section for all-hadronic decay modes of 1.7 pb is obtained. Hence calculate a value for the W mass.

Dynamic quarks

9

In this chapter, we consider the structure of hadrons and discuss the most direct dynamical evidence that hadrons are made from quarks. This evidence comes from the scattering of high-energy leptons off protons and neutrons. The relatively high cross section for reactions at large transverse momentum, so-called deep inelastic scattering (DIS), leads to the conclusion that these hadrons are made up of point-like constituents. To observe the structure within hadrons, a resolution smaller than the size of the hadron is required; hence the wavelength of the probe should satisfy $\lambda \ll R$ (where R is the radius of a hadron). From the de Broglie relationship¹ $\lambda = h/p$, therefore, high-energy experiments are required to study this structure. Using the uncertainty relation in a form more suitable for high-energy physics, $\Delta x \Delta p c \leq \hbar c$, implies that we require an energy transfer of the order of at least 20 GeV to achieve a resolution of 10^{-2} fm.²

Other evidence that hadrons are made of quarks was discussed in Chapter 5, where the static quark model of hadrons was used to explain the observed multiplets of hadrons and their masses and magnetic moments. In Section 9.1, we will consider scattering off a nucleus and in Section 9.2 scattering off individual nucleons. Then we will discuss how the DIS data can be explained in terms of the quark–parton model (QPM). The DIS data give indirect evidence for the existence of gluons, and we will consider more direct evidence for gluons from e^+e^- experiments. A brief discussion of QCD will be given, and this will be used to explain the success of the naive QPM. Finally, the QPM will be extended to hadron–hadron collisions.

9.1 Rutherford scattering	235
9.2 Scattering from nucleons	239
9.3 Quark–parton model	240
9.4 Neutrino interactions	242
9.5 Charged-lepton probes	251
9.6 QCD introduction	259
9.7 Hadron–hadron collisions	272
Chapter summary	274
Further reading	275
Exercises	275

¹Count Louis de Broglie 1892–1987.

² $\hbar c = 197$ MeV fm.

9.1 Rutherford scattering

The prototype for all scattering experiments is the Rutherford scattering experiment that discovered the atomic nucleus. The experiment involved scattering α particles off a gold foil and measuring the angular distribution of the scattered particles. We can calculate the transition rate from Fermi’s Golden Rule, which gives

$$w_{fi} = 2\pi |\langle f | H' | i \rangle|^2 \rho(E) \quad (9.1)$$

where H' is the Hamiltonian for the perturbation that causes the transition between initial state i and final state f and $\rho(E)$ is the density-of-states factor. In this case, the perturbation is given by the Coulomb

interaction between the α particle (atomic number Z_1) and the nucleus (atomic number Z_2). If all the positive charge is contained in a point-like nucleus, we can write the potential as

$$V(r) = \frac{Z_1 Z_2 \alpha}{r} \quad (9.2)$$

where $\alpha = e^2/(4\pi\epsilon_0)$ is the fine-structure constant. The matrix element is then given by the integral of $V(r)$ between the initial and final states. We can use plane waves for the initial and final states of the α particles: $\Psi_i = \exp(i\mathbf{k}_i \cdot \mathbf{r})$ and $\Psi_f = \exp(i\mathbf{k}_f \cdot \mathbf{r})$. This gives

$$\langle f|H'|i\rangle = \int \exp(i\mathbf{k}_i \cdot \mathbf{r})V(r) \exp(-i\mathbf{k}_f \cdot \mathbf{r}) d^3\mathbf{r} \quad (9.3)$$

Substituting for $V(r)$ from eqn 9.2 into eqn 9.3 and then using $\mathbf{q} = \mathbf{k}_i - \mathbf{k}_f$ for the momentum transfer gives

$$\langle f|H'|i\rangle = Z_1 Z_2 \alpha \int \frac{e^{i\mathbf{q} \cdot \mathbf{r}}}{r} d^3\mathbf{r} \quad (9.4)$$

It is convenient to use spherical polar coordinates and to take the z axis of \mathbf{r} to lie along \mathbf{q} so that $\mathbf{q} \cdot \mathbf{r} = qr \cos \theta$. Then we can perform the integral over ϕ and θ :

$$\langle f|H'|i\rangle = Z_1 Z_2 2\pi\alpha \int e^{iqr \cos \theta} r dr d \cos \theta \quad (9.5)$$

$$= \frac{Z_1 Z_2 2\pi\alpha}{iq} \int_0^\infty (e^{iqr} - e^{-iqr}) dr \quad (9.6)$$

This integral is divergent, which is related to the fact that the scattering cross section for scattering of two particles via an inverse-square force law is infinite. We will proceed by modifying the potential by a factor $e^{-r/a}$ and at the end of the calculation we will let $a \rightarrow \infty$. Then we can perform the integral in eqn 9.6 to give

$$\langle f|H'|i\rangle = Z_1 Z_2 \frac{2\pi\alpha}{iq} \left(\frac{1}{1/a - iq} - \frac{1}{1/a + iq} \right) \quad (9.7)$$

and then, if we let $a \rightarrow \infty$, we obtain the matrix element

$$\langle f|H'|i\rangle = \frac{4\pi Z_1 Z_2 \alpha}{q^2} \quad (9.8)$$

The density of states for a unit volume V (this volume is arbitrary and will cancel with the flux in the calculation of the cross section) is given by (see Chapter 2)

$$dN_f = \frac{d^3\mathbf{p}_f}{(2\pi)^3} \quad (9.9)$$

We make a change of variable using

$$\frac{dp}{dE} = \frac{1}{v_f} \quad (9.10)$$

where v_f is the final velocity. As we are assuming an infinitely massive nucleus, we can neglect the nuclear recoil energy. Writing $d^3p_f = p_f^2 dp_f d\Omega$ and substituting into eqn 9.9 gives

$$dN_f = \frac{p_f^2 dp_f d\Omega}{(2\pi)^3} \quad (9.11)$$

The differential scattering cross section is given by R/v , where v is the velocity of the incident beam relative to the scattering centre and R is the reaction rate.³

Substituting from eqns 9.8, 9.9, and 9.11 into eqn 9.1 gives the differential cross section as (see Exercise 9.1)

$$\frac{d\sigma}{d\Omega} = \frac{4(Z_1 Z_2)^2 \alpha^2 (m_e)^2}{q^4} \quad (9.12)$$

This is an example of a scaling cross section, since it does not depend on any fixed scale in the problem. This arises because we have assumed that the nucleus is a point charge. If instead we allow for a charge density distribution $\rho(r)$, then the potential becomes

$$V(\mathbf{r}) = Z\alpha \int \frac{\rho(R)}{|\mathbf{r} - \mathbf{R}|} d^3\mathbf{R} \quad (9.13)$$

Then the matrix element is modified from eqn 9.8 to become

$$\langle f|H'|i\rangle = Z\alpha \int d^3\mathbf{R} \int e^{i\mathbf{q}\cdot\mathbf{r}} \frac{\rho(R)}{|\mathbf{r} - \mathbf{R}|} d^3\mathbf{r} \quad (9.14)$$

Letting $\mathbf{s} = \mathbf{r} - \mathbf{R}$, we can then write the matrix element as

$$\langle f|H'|i\rangle = Z\alpha \iint e^{i\mathbf{q}\cdot\mathbf{R}} \rho(R) d^3\mathbf{R} \frac{e^{i\mathbf{q}\cdot\mathbf{s}}}{|\mathbf{s}|} d^3\mathbf{s} \quad (9.15)$$

Hence the matrix element is modified by a multiplicative factor called the 'form factor'

$$F(q^2) = \int e^{i\mathbf{q}\cdot\mathbf{R}} \rho(R) d^3\mathbf{R} \quad (9.16)$$

The form factor can be seen to be the Fourier transform of the charge distribution into momentum space. By measuring the deviations from the pure Rutherford scattering cross section, eqn 9.12, the form factor can be determined. Hence the mean size and charge density distribution of the nucleus can be inferred. However, to obtain any meaningful data on the nuclear size, one must have data with momentum transfer large enough to satisfy the inequality $qR_{\text{nucleus}} \gg 1$, where R_{nucleus} is

³We have set the normalization volume to be unity, and therefore the incident flux is $1/v$, and we have assumed a single target nucleus.

the mean size of the nucleus. The cross section no longer shows scaling behaviour, because of the effect of the finite size of the nucleus. For example, if the nucleus had an exponential charge distribution

$$\rho(\mathbf{r}) = \rho_0 e^{-mr} \tag{9.17}$$

then the form factor would be given by

$$F(q^2) = \frac{8\pi m \rho_0}{(q^2 + m^2)^2} \tag{9.18}$$

Note that for $q \ll m$, $F(q^2)$ is a constant, whereas for $q \gg m$, $F(q^2) \sim q^{-4}$, i.e. the cross section is suppressed by a factor of $1/q^8$. This feature that the cross section is suppressed for values of q large compared with $1/R$, where R is the size of the nucleus, is a general result and will be true whatever the precise form of the charge distribution. An example of some real data [53] and a fit to the charge density distribution are shown in Fig. 9.1.

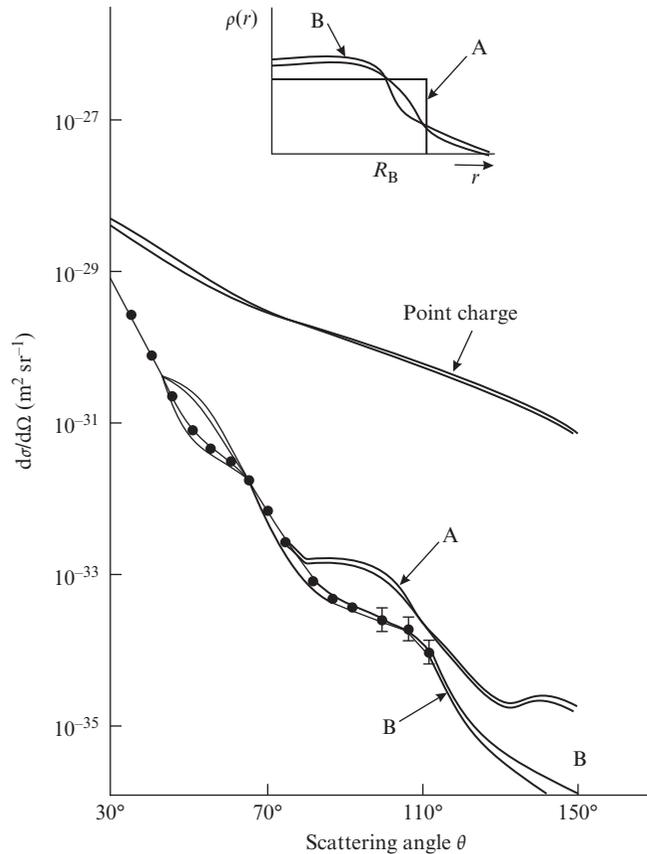


Fig. 9.1 Cross section for the scattering of 153 MeV electrons off gold nuclei, with fits to the nuclear charge distribution for a sharp edge (A) and a realistic rounded edge (B). From [53].

9.2 Scattering from nucleons

In the previous section, we have seen how deviations from Rutherford scattering can be used to measure the charge distribution of the nucleus. Now we will consider the problem of how to look for evidence of individual nucleons (protons and neutrons) in a nucleus. The answer is to do scattering experiments where we look for evidence of inelastic scattering off the whole nucleus and elastic scattering off a single nucleon. The kinematics are defined in Fig. 9.2.

The invariant mass W of the recoiling nucleon of mass M is given by

$$W^2 = (E^*)^2 - |\mathbf{p}^*|^2 \quad (9.19)$$

and the 4-momentum transfer q is found from 4-momentum conservation at the lower vertex as $q = (E^* - M, \mathbf{p}^*)$. Therefore, we can write down the square of the 4-momentum transfer:

$$\begin{aligned} q^2 &= (E^* - M)^2 - |\mathbf{p}^*|^2 \\ &= (E^*)^2 - |\mathbf{p}^*|^2 - 2ME^* + M^2 \end{aligned} \quad (9.20)$$

Substituting for W from eqn 9.19 gives

$$q^2 = W^2 - 2ME^* + M^2 \quad (9.21)$$

From energy conservation, $\nu = E - E' = E^* - M$ and therefore, from eqn 9.21,

$$q^2 = W^2 - M^2 - 2M\nu \quad (9.22)$$

For elastic scattering off the entire nucleus, the mass of the nucleus is unchanged in the collision, and hence $W = M$, so we get a peak in the change in the electron energy given by

$$\nu = -\frac{q^2}{2M_{\text{nucleus}}} \quad (9.23)$$

Similarly if we had scattering off a single nucleon, then we would expect

$$\nu = -\frac{q^2}{2M_{\text{nucleon}}} \quad (9.24)$$

Some example data [87] are shown in Fig. 9.3, where the cross section for electron scattering off helium is plotted as a function of the scattered electron energy E' . A sharp peak is seen at the value expected for elastic scattering off the entire nucleus and a smeared peak is seen for elastic scattering off individual nucleons. The smearing is caused by the Fermi motion of the nucleons within the nucleus. Naively, we might hope that we could see evidence for the quark structure of the nucleons in a similar way, but now the Fermi motion is so big that it completely smears out the peak corresponding to elastic scattering off a quark.

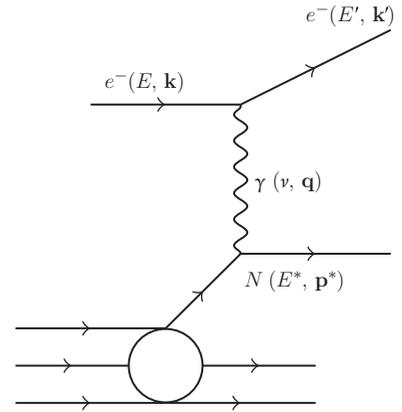


Fig. 9.2 Kinematics for scattering off individual nucleons. We assume that the nucleus is initially at rest in the laboratory frame of reference.

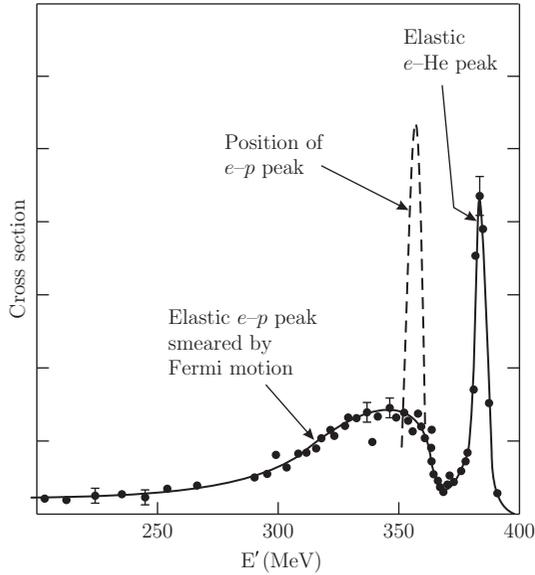


Fig. 9.3 Scattering of 400 MeV electrons off ${}^4\text{He}$, with the scattering angle fixed at $\theta = 45^\circ$. E' is the energy of the scattered electron [87].

9.3 Quark-parton model

This section introduces the quark-parton model (QPM), which will be used in the following sections to make predictions for deep inelastic scattering processes. Our approach is not historically accurate—it took physicists a long time to take the experimental data seriously enough to really believe in quarks. However, it is much easier to understand. We will compare the predictions with the experimental data and show that the key features are confirmed by the data, namely that the nucleons contain spin- $\frac{1}{2}$, point-like particles with the fractional electric charges expected in the quark model.

In the QPM, we assume that the inelastic scattering of a lepton at large momentum transfer q with a nucleon is due to the elastic scattering off a quark via exchange of a virtual boson (γ , W , or Z). First, we will review the kinematics of the reaction in Section 9.3.1; then in Sections 9.4 and 9.5, we develop the dynamics to allow us to predict the form of the differential cross sections. We then compare QPM predictions with the experimental data and discuss many internal consistency checks that give us confidence in the model.

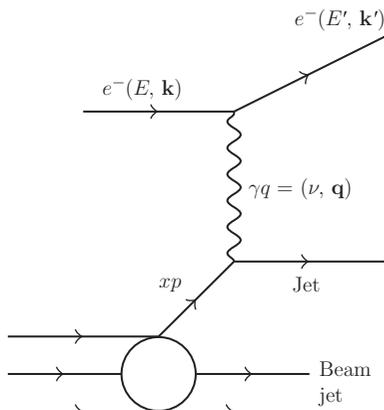


Fig. 9.4 Kinematics for quark-parton scattering.

9.3.1 Kinematics of deep inelastic scattering

The key idea of the QPM is shown in Fig. 9.4, which also defines the kinematics. We assume that the quarks each have a mass that is a fraction x of the nucleon mass; i.e. the quark mass $m = xM$. The energy of the struck quark after the scattering is $E^* = \nu + xM$, where ν is energy transferred by the virtual photon. In terms of the incoming and

outgoing electron energies, $\nu = E - E'$, and the 3-momentum transfer is $\mathbf{p}^* = \mathbf{q}$. Then, using $E^2 - p^2 = m^2$ for the struck quark after the elastic collision, we get

$$(\nu + xM)^2 - |\mathbf{q}|^2 = (xM)^2 \quad (9.25)$$

Multiplying this out gives

$$\nu^2 + 2xM\nu + (xM)^2 - |\mathbf{q}|^2 = (xM)^2 \quad (9.26)$$

Defining $Q^2 = -q^2 = -(\nu^2 - |\mathbf{q}|^2)$, we solve for the parton mass fraction x as

$$x = \frac{Q^2}{2M\nu} \quad (9.27)$$

Note that x can be determined purely from a measurement of the scattered lepton momentum. This is important from an experimental point of view, since it is generally easier to make precise measurements of the momenta of electrons and muons than those of hadrons.⁴

The variable x can also be considered as representing the fraction of the nucleon momentum carried by a parton. In DIS, the 4-momentum transfer q is large enough for the mass of the quark to be neglected in comparison with its energy. Hence, considering the 4-momentum of the quark after the scattering and using the fact that the quark mass is negligible before the interaction,

$$(p_{\text{quark}} + q)^2 = 0$$

that is,

$$p_{\text{quark}}^2 + 2p_{\text{quark}} \cdot q + q^2 = 0 \quad (9.28)$$

Letting $p_{\text{quark}} = xp$, we can solve for x from eqn 9.28 as

$$x = \frac{Q^2}{2p \cdot q} \quad (9.29)$$

Now, $p \cdot q$ is a Lorentz invariant, so we can evaluate it in any frame we wish. Consider the nucleon rest frame, in which $p = (M, 0)$, and use eqn 9.29:

$$x = \frac{Q^2}{2M\nu} \quad (9.30)$$

This value of x is the same as that obtained from eqn 9.27, so we can consider the variable x either as representing the mass fraction of the nucleon carried by a quark or as the momentum fraction of the nucleon carried by the quark.⁵

Before we can start the discussion of the dynamics of the scattering process, we need to evaluate one more piece of kinematics, namely the relation between the laboratory energies and the CMS scattering angle θ^* .

⁴Early experiments used ‘single-arm spectrometers’, which could only measure the scattered lepton momentum and not that of the hadronic jet.

⁵Strictly speaking, this is only correct in the ‘infinite-momentum frame’, in which the nucleus has an infinite momentum along the collision axis so that we can neglect transverse components of the momentum.

The Lorentz transformation from the CMS to the laboratory system for the scattered and initial leptons (we assume $\beta = 1$) is

$$\begin{aligned} E'_{\text{lab}} &= \gamma E^*(1 + \cos \theta^*) \\ E_{\text{lab}} &= \gamma E^*(1 + 1) \end{aligned} \quad (9.31)$$

We define the scaled fractional energy transfer for the scattered lepton to be $y = \nu/E$, so $0 < y < 1$. Then we can relate the CMS scattering angle θ^* and y as

$$y = \frac{1}{2}(1 - \cos \theta^*)^6 \quad (9.32)$$

⁶Note that the variable y is only defined in the lab frame.

Having completed our discussion of the kinematics of lepton–nucleon scattering, we can now look at the dynamics. We will start with neutrino probes in Section 9.4.1, because the spin structure is simpler for this case than for the case of charged lepton probes, which we will consider in Section 9.5.

9.4 Neutrino interactions

We are now ready to use the QPM to explain the DIS data for neutrino probes.⁷ We will proceed in the following steps:

⁷We will consider electron neutrinos in the following discussion, but in the DIS regime the masses of the leptons are negligible, so the same results hold for muon neutrinos.

- (1) Calculate the elastic scattering cross sections for
 - (a) $\bar{\nu}_e e \rightarrow \bar{\nu}_e e$
 - (b) $\nu_e e \rightarrow \nu_e e$
- (2) Generalize to the elastic scattering of (anti)neutrinos on (anti)quarks.
- (3) Generalize to the case of (anti)neutrinos scattering off nucleons consisting of quarks and antiquarks.

As discussed in Chapter 7, weak interactions are parity-violating; the angular distribution in β decay is maximally parity-violating and the leptons are emitted polarized. The parity violation is a V–A interaction, where V and A stand for vector and axial vector couplings, respectively.⁸

⁸In principle, there are five possible couplings, each with a different experimental signature. See Further Reading to follow this up.

We consider for now only virtual W exchange (i.e. we neglect Z exchange) and so we have a pure V–A interaction. In the high-energy limit, this leads to a very simple spin structure for the process whereby the W bosons couple only to negative-helicity leptons and positive-helicity antileptons, where the helicity is defined by the normalized projection of the spin \mathbf{s} onto the momentum \mathbf{p} of the particle:

$$H = \frac{\mathbf{p} \cdot \mathbf{s}}{|\mathbf{p}| |\mathbf{s}|} \quad (9.33)$$

so that for spin- $\frac{1}{2}$ fermions the eigenvalues of helicity are $+1$ and -1 . Therefore, the spin structures for the interactions $\bar{\nu}_e e \rightarrow \bar{\nu}_e e$ and

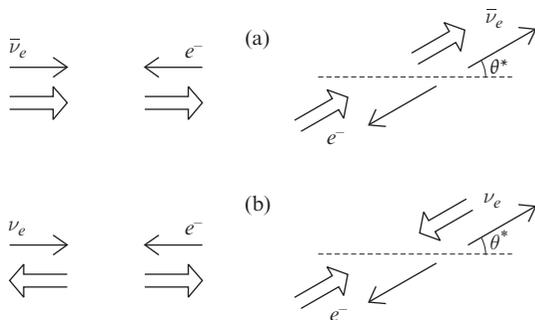


Fig. 9.5 Spin structure for (a) $\bar{\nu}_e e$ scattering and (b) $\nu_e e$ scattering. The left (right) diagram is before (after) the scatter.

$\nu_e e \rightarrow \nu_e e$ are as shown in Fig. 9.5. For the interaction in Fig. 9.5(a), the overall initial state must have an angular momentum about the z axis (the initial direction of the $\bar{\nu}_e$) given by the quantum number $J_z = 1$. The interaction proceeds via the creation of a virtual W . The spin of a real W is measured to be 1 and we will assume that the spin of the virtual W is also 1. Hence the initial state must have $J = 1, J_z = 1$ and by conservation of angular momentum the final state will have the same angular momentum quantum numbers. However, we know that the scattered leptons will be in eigenstates of helicity (positive for the $\bar{\nu}_e$ and negative for the e^-); i.e. the $\bar{\nu}_e$ will have its spin along its direction of motion. Hence, considering a z' axis that is rotated from the z axis to lie along the final $\bar{\nu}_e$ direction of motion, $J_{z'} = \frac{1}{2}$. Similarly, the electron has $J_{z'} = \frac{1}{2}$. Therefore, the amplitude for the reaction in Fig. 9.5(a) can be found by projecting out the spin states onto the rotated axes. This requires the rotation matrix for spin- $\frac{1}{2}$ particles (see Chapter 2 and Exercise 2.4):

$$d_{m'm}^j = \begin{pmatrix} \cos \frac{1}{2}\theta & -\sin \frac{1}{2}\theta \\ \sin \frac{1}{2}\theta & \cos \frac{1}{2}\theta \end{pmatrix} \quad (9.34)$$

The elements of this matrix give the amplitude for a state with a given quantum number m to be found with a value of m' after a rotation by a polar angle θ .

9.4.1 Cross section for neutrino–electron elastic scattering

We are now in a position to calculate the cross section for elastic $\nu_e e$ and $\bar{\nu}_e e$ scattering. From the spin structure for the reaction $\bar{\nu}_e e \rightarrow \bar{\nu}_e e$ shown in Fig. 9.5(a), the amplitude as a function of polar angle can be calculated by projecting the spin states onto the rotated axes. This projection changes the electron ($\bar{\nu}_e$) state from $m = \frac{1}{2}$ to $m' = \frac{1}{2}$, and hence the amplitude as a function of polar angle is given by⁹

$$A(\theta) = (d_{1/2,1/2}^{1/2}(\theta))^2 \quad (9.35)$$

⁹We have chosen to apply the spin- $\frac{1}{2}$ rotation matrices to the $\bar{\nu}_e$ and e states separately. The same result could have been obtained by considering the combined system and using spin-1 rotation matrices.

and, substituting for the appropriate element of the rotation matrix from eqn 9.34, we get

$$A(\theta) = \left(\cos \frac{1}{2}\theta \right)^2 = \frac{1}{2}(1 + \cos \theta) \tag{9.36}$$

The cross section is proportional to $|A(\theta)|^2$. We can use eqn 9.32 to change variables from θ to y . From the two-body phase space, the cross section should also be proportional to $(p^*)^2$, where p^* is the CMS momentum. We define s to be the CMS energy squared, and then¹⁰ $(p^*)^2 = s/4$. The amplitude of the charged-current weak interaction is proportional to the Fermi coupling constant G_F . Hence, from eqn 9.36 and putting in the correct numerical factor,¹¹ we get the differential cross section for elastic $\bar{\nu}_e e$ scattering:

$$\frac{d\sigma(\bar{\nu}_e e \rightarrow \bar{\nu}_e e)}{dy} = \frac{G_F^2 s}{\pi} (1 - y)^2 \tag{9.37}$$

Similarly, for the elastic scattering reaction $\nu_e e$, from the spin structure shown in Fig. 9.5(b), the angular momentum of the initial state must have $J_Z = 0$. The possible values of the total angular momentum quantum number J are therefore 0 or 1.

We can calculate the amplitude for the reaction $\nu_e e \rightarrow \nu_e e$, given the amplitude for the reaction $\bar{\nu}_e e \rightarrow \bar{\nu}_e e$. From the solutions to the Dirac equation, we saw that we can formally represent the states that apparently have negative energy as positive-energy states travelling backwards in time. Hence we can use crossing symmetry to relate the amplitudes of reactions with a particle exchanged for its antiparticle. This states that for two diagrams related by crossing, the structures of the matrix elements are the same and we have only to replace the momentum of the incoming (outgoing) particles with minus that of the outgoing (incoming) antiparticles. From Fig. 9.6, we can rewrite the amplitude for $\bar{\nu}_e e$ scattering (eqn 9.36) in terms of the 3-momenta and the magnitude of the CMS momentum of the particles, p^* , as

$$A_{\bar{\nu}}(\theta) = \frac{(p^*)^2 + \mathbf{p}_1 \cdot \mathbf{p}_3}{2(p^*)^2} \tag{9.38}$$

and, in the CMS, $\mathbf{p}_3 = -\mathbf{p}_4$, so we can rewrite this as

$$A_{\bar{\nu}}(\theta) = \frac{(p^*)^2 - \mathbf{p}_1 \cdot \mathbf{p}_4}{2(p^*)^2} \tag{9.39}$$

We can rewrite eqn 9.39 in terms of 4-vectors as

$$A_{\bar{\nu}}(\theta) = \frac{p_1 \cdot p_4}{2(p^*)^2} \tag{9.40}$$

and then use crossing symmetry to find the amplitude for ν_e scattering by the substitution $p_1 \leftrightarrow -p_3$:

$$A_{\nu}(\theta^*) = -\frac{p_3 \cdot p_4}{(2p^*)^2} \tag{9.41}$$

¹⁰We can neglect the electron mass since $p^* \gg m_e$.

¹¹The calculation of the correct numerical factors requires an application of the Feynman rules. This is described in the standard graduate-level textbooks—see Griffiths in Further Reading for an example.

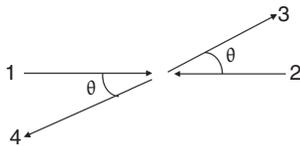


Fig. 9.6 Definition of the angles used in the crossing symmetry argument.

We use $(p_3 + p_4)^2 = p_3^2 + p_4^2 + 2p_3 \cdot p_4$ and we can neglect the masses. Therefore, $p_3 \cdot p_4 = (2p^*)^2$, and we can see from eqn 9.41 that the amplitude for $\nu_e e$ scattering is isotropic. Hence, putting in the factors of G_F , s , and overall normalization as for $\bar{\nu}_e e \rightarrow \bar{\nu}_e e$, we get the differential cross section for $\nu_e e$ elastic scattering as

$$\frac{d\sigma(\nu_e e \rightarrow \nu_e e)}{dy} = \frac{G_F^2 s}{\pi} \quad (9.42)$$

In summary, the $\bar{\nu}_e e^-$ scattering cross section has a factor of $(1 - y)^2$ whereas the $\nu_e e^-$ cross section has a factor of 1. A similar argument can be used to obtain the cross section for $\bar{\nu}_e e^+$ from that for $\bar{\nu}_e e^-$, and we find that $\bar{\nu}_e e^+$ scattering has a factor of 1 whereas $\nu_e e^+$ scattering has a factor of $(1 - y)^2$.

9.4.2 Neutrino–quark scattering

The next step is to generalize the results for neutrino–electron scattering to the case of neutrino–quark scattering. The universality of the charged-current weak interactions means that the weak interaction has the same strength for quarks as for electrons (see Chapter 7). Unfortunately, we do not have free quarks available as targets, so we have to use the quarks that are confined in nucleons. One of the main techniques when doing dynamic quark modelling is to consider the kinematics of elastic scattering of a neutrino with a quark that carries a fraction x of the nucleon momentum (see Fig. 9.7).

The Lorentz-invariant CMS energy squared of the neutrino–quark system is given by

$$\begin{aligned} \hat{s} &= (p^* + xp^*)^2 - (p^* - xp^*)^2 \\ &= 4x(p^*)^2 = xs \end{aligned} \quad (9.43)$$

Hence we can write down the cross sections for $\nu_e(\bar{\nu}_e)$ elastic scattering on $q(\bar{q})$ by analogy with the $\nu_e(\bar{\nu}_e)$ elastic scattering cross sections¹² (eqns 9.37 and 9.42) by the substitution¹³ $s \rightarrow \hat{s} = xs$:

$$\begin{aligned} \frac{d\sigma(\nu_e q \rightarrow \nu_e q)}{dy} &= \frac{G_F^2 sx}{\pi} \\ \frac{d\sigma(\bar{\nu}_e q \rightarrow \bar{\nu}_e q)}{dy} &= \frac{G_F^2 sx}{\pi} (1 - y)^2 \\ \frac{d\sigma(\nu_e \bar{q} \rightarrow \nu_e \bar{q})}{dy} &= \frac{G_F^2 sx}{\pi} (1 - y)^2 \\ \frac{d\sigma(\bar{\nu}_e \bar{q} \rightarrow \bar{\nu}_e \bar{q})}{dy} &= \frac{G_F^2 sx}{\pi} \end{aligned} \quad (9.44)$$

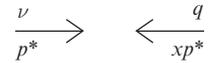


Fig. 9.7 Neutrino–quark scattering kinematics.

¹²We ignore threshold effects, which can be important for the case of charm production near threshold.

¹³The variable y is unchanged here because it depends only on the neutrino quantities, not on those of the quark.

9.4.3 Neutrino–nucleon cross sections

We can now put together all the pieces and calculate the cross section for neutrino–nucleon scattering in the DIS regime. The basic assumption of the QPM is that the cross section for scattering leptons off nucleons is given by the incoherent sum of scattering off free quarks. This assumption is only valid for large values of Q^2 ($Q^2 \gg 1 \text{ GeV}^2$). We will discuss the justification for this assumption when we consider the theory of strong interactions in Section 9.6.3. Let $q(x)$ be the probability distribution function of quarks with a momentum fraction x ; i.e. the probability of finding a quark with momentum in the range $(x, x + dx)$ is $q(x) dx$. Similarly, let $\bar{q}(x)$ be the equivalent distribution for antiquarks. The QPM makes no prediction for $q(x)$ or $\bar{q}(x)$, so we have to obtain $q(x)$ from fits to experimental data. Nevertheless, the QPM does make many clear predictions that can be tested experimentally. With the above assumptions, we are finally able to write down the cross sections for neutrino and antineutrino scattering off nucleons as

$$\frac{d^2\sigma(\nu_e N)}{dx dy} = \frac{G_F^2 sx}{\pi} [q(x) + (1-y)^2 \bar{q}(x)] \quad (9.45)$$

$$\frac{d^2\sigma(\bar{\nu}_e N)}{dx dy} = \frac{G_F^2 sx}{\pi} [(1-y)^2 q(x) + \bar{q}(x)] \quad (9.46)$$

We can compare these predictions with the measured y distributions for neutrino and antineutrino beams [115] shown in Fig. 9.8.

If the nucleons contained only quarks and not antiquarks, then we would expect the neutrino data to be constant in y and for the antineutrino data to show a $(1-y)^2$ dependence. The data can be fit by a mixture of constant and $(1-y)^2$ components, which tells us that nucleons are composed of quarks and antiquarks. The proportion of quarks to antiquarks can thus be estimated from these data. The fact that the data do

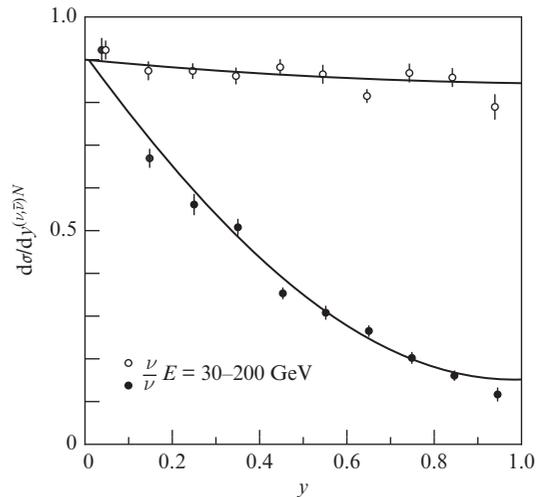


Fig. 9.8 Neutrino and antineutrino cross sections on nuclei as functions of y . From [71].

fit the QPM prediction for the y distribution is a very important test of the theory. It tells us that the neutrinos are interacting with free spin- $\frac{1}{2}$ particles (the quarks and antiquarks) according to the parity-violating $V-A$ interaction. We can estimate the total cross sections by integrating the differential cross sections in eqns 9.45 and 9.46. The y integral is trivial to perform and the x integral gives a constant of the order of unity (depending on the unknown $q(x)$ distribution). If we assume that the \bar{q} content of the nucleon is negligible, we get

$$\begin{aligned}\sigma(\nu N) &\approx G_F^2 s \\ \sigma(\bar{\nu} N) &\approx G_F^2 s/3\end{aligned}\tag{9.47}$$

This prediction is compared with the experimental data [115] in Fig. 9.9.¹⁴ The neutrino cross sections are found to be larger than the antineutrino cross sections, as expected if the nucleons consist mainly of quarks as opposed to antiquarks—but not by as much as a factor of 3, which indicates that there are some antiquarks in protons. More significantly, the fact that the cross section scales like s is telling us that the quarks are behaving as point-like particles. If the quarks had a finite size, then the cross section would be suppressed by a form factor.

¹⁴For a fixed target nucleon, $s \approx 2m_N E_\nu$, where E_ν is the neutrino beam energy.

9.4.4 Parton distribution functions

In the previous section, we have seen that the QPM can explain some of the key features of neutrino–nucleon interactions. However, the quark distribution functions are not predicted by the model and have to be determined from experimental data. In this section, we will explain how this is done and discuss some useful consistency checks of the theory.

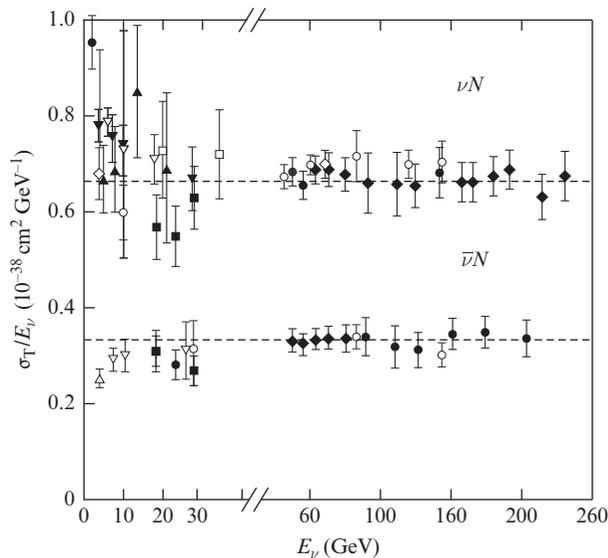


Fig. 9.9 Neutrino and antineutrino cross sections. From [115].

First, it is convenient to rewrite the ν nucleon cross sections given by eqn 9.46 in a form more suitable for comparison with the experimental data. We use the relationship

$$q(x) + (1-y)^2 \bar{q}(x) = [q(x) + \bar{q}(x)] \frac{1 + (1-y)^2}{2} + [q(x) - \bar{q}(x)] \frac{1 - (1-y)^2}{2} \quad (9.48)$$

and substitute into eqn 9.46 to obtain

$$\frac{d^2 \sigma(\nu N)}{dx dy} = \frac{G_F^2 s x}{\pi} \left\{ [q(x) + \bar{q}(x)] \frac{1 + (1-y)^2}{2} \pm [q(x) - \bar{q}(x)] \frac{1 - (1-y)^2}{2} \right\} \quad (9.49)$$

We can compare the cross-section formula of eqn 9.49 with a general phenomenological formula that is allowed by Lorentz invariance:¹⁵

$$\frac{d^2 \sigma(\nu N)}{dx dy} = \frac{G_F^2 s x}{2\pi} \left[\frac{F_2(x, q^2)}{x} \frac{1 + (1-y)^2}{2} \pm F_3(x, q^2) \frac{1 - (1-y)^2}{2} \right] \quad (9.50)$$

¹⁵This is neglecting the ‘longitudinal’ structure function F_L , which is expected to be a good approximation at large values of momentum transfer Q^2 . Note also that some older analyses of DIS data use the structure function F_L rather than F_L , where $F_L = F_2 - 2xF_1$, but this hides the evidence that F_L is small.

¹⁶We will return to this later in the chapter.

The so-called structure functions F_2 and F_3 can be determined by fits to the experimental data. Note that in general the functions F_i can depend on q^2 as well as x . The QPM does not predict any dependence on q^2 and to a first approximation this agrees with data.¹⁶ Then, from a comparison of eqns 9.49 and 9.50, we can obtain the quark probability distribution functions as

$$\begin{aligned} \sum_i q_i(x) &= \frac{1}{4} \left[\frac{F_2(x)}{x} + F_3(x) \right] \\ \sum_i \bar{q}_i(x) &= \frac{1}{4} \left[\frac{F_2(x)}{x} - F_3(x) \right] \end{aligned} \quad (9.51)$$

where the sum runs over the different quark flavours. Or, equivalently, we can write the inverse relation to give the structure functions in terms of the quark distribution functions:

$$\begin{aligned} F_2(x) &= 2 \sum_i x [q_i(x) + \bar{q}_i(x)] \\ F_3(x) &= 2 \sum_i [q_i(x) - \bar{q}_i(x)] \end{aligned} \quad (9.52)$$

So far in the discussion, we have not defined precisely which quark flavours we are considering for the quark distribution functions. We assume that the nucleons contain only the light u , d , and s quarks and the corresponding antiquarks (i.e. we neglect the heavy-quark c , b , and t content).

The Feynman diagrams corresponding to the interactions of neutrinos and antineutrinos on these light quarks are shown in Fig. 9.10, from which we see that neutrinos interact with d and s quarks and \bar{u} antiquarks, whereas antineutrinos interact with u quarks and \bar{d} and \bar{s} antiquarks. It is conventional to define the quark distribution functions as referring to protons. (It should be noted that, at this stage, we are ignoring the differences between the weak eigenstates and the mass eigenstates for the quarks. This is explained in Chapter 7.)

We assume that the s and \bar{s} quark distribution functions are the same in neutrons and protons. Therefore, using eqn 9.52, we can write the structure functions for proton targets in terms of the quark distribution functions as

$$\begin{aligned} F_2^{\nu p}(x) &= 2x[d(x) + \bar{u}(x) + s(x)] \\ F_2^{\bar{\nu} p}(x) &= 2x[u(x) + \bar{d}(x) + \bar{s}(x)] \\ F_3^{\nu p}(x) &= 2[d(x) + s(x) - \bar{u}(x)] \\ F_3^{\bar{\nu} p}(x) &= 2[u(x) - \bar{s}(x) - \bar{d}(x)] \end{aligned} \quad (9.53)$$

For a neutron target, we will assume isospin (SU(2) flavour) symmetry; i.e. we assume that the d quarks (\bar{d} antiquarks) in a neutron have the same distribution function as the u quarks (\bar{u} antiquarks) in a proton. Then we can simply use the substitution $u \leftrightarrow d$ and $\bar{u} \leftrightarrow \bar{d}$ in eqn 9.53 to write down the neutron structure functions in terms of the quark distribution functions as

$$\begin{aligned} F_2^{\nu n}(x) &= 2x[u(x) + \bar{d}(x) + s(x)] \\ F_2^{\bar{\nu} n}(x) &= 2x[d(x) + \bar{u}(x) + s(x)] \\ F_3^{\nu n}(x) &= 2[u(x) + s(x) - \bar{d}(x)] \\ F_3^{\bar{\nu} n}(x) &= 2[d(x) - \bar{s}(x) - \bar{u}(x)] \end{aligned} \quad (9.54)$$

For an ‘isoscalar’ target, i.e. one with equal numbers of neutrons and protons in an isospin $I = 0$ state,¹⁷ we can write down the structure functions from the average of the proton and neutron structure functions (eqns 9.53 and 9.54). We also assume $s(x) = \bar{s}(x)$, because s and \bar{s} quarks and antiquarks have to be created together by the strong interaction, which conserves strangeness. We have

$$\begin{aligned} F_2^{\nu N}(x) &= x[u(x) + d(x) + \bar{u}(x) + \bar{d}(x) + 2s(x)] \\ F_2^{\bar{\nu} N}(x) &= x[u(x) + d(x) + \bar{u}(x) + \bar{d}(x) - 2s(x)] \\ F_3^{\nu N} &= u(x) - \bar{u}(x) + d(x) - \bar{d}(x) + 2s(x) \\ F_3^{\bar{\nu} N} &= u(x) - \bar{u}(x) + d(x) - \bar{d}(x) - 2s(x) \end{aligned} \quad (9.55)$$

It is convenient to divide the quark distribution functions into ‘valence’ and ‘sea’ parts:

$$u(x) = u_v(x) + u_s(x) \quad (9.56)$$

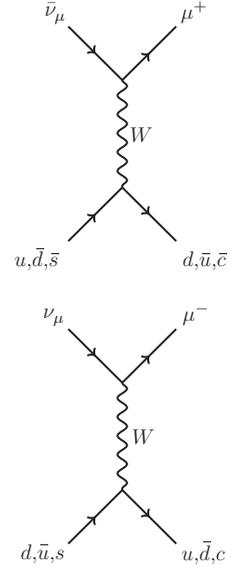


Fig. 9.10 Antineutrino and neutrino scattering off different flavours of quarks.

¹⁷As is the case for many easy-to-use targets in scattering experiments, e.g. carbon.

where, by definition, $u_s(x) = \bar{u}(x)$. Then, if we consider an average of neutrino and antineutrino data, we have

$$F_3(x) = u(x) - \bar{u}(x) + d(x) - \bar{d}(x) \tag{9.57}$$

Hence, by the definition of valence and sea quarks, we expect

$$F_3(x) = u_v(x) + d_v(x) \tag{9.58}$$

Since there are three valence quarks in a proton, we can integrate $F_3(x)$ to get the Gross–Llewellyn Smith (GLS) sum rule [81]

$$S_{\text{GLS}} = \int_0^1 F_3(x) dx = \int_0^1 [u_v(x) + d_v(x)] dx = 3 \tag{9.59}$$

This prediction of the QPM is compared with experimental data in Fig. 9.11. The measured value for the GLS sum [98] is $S_{\text{GLS}} = 2.50 \pm 0.018(\text{statistical}) \pm 0.078(\text{systematic})$, which is slightly lower than the predicted value of 3 in the simple QPM. However, the difference can be understood in terms of higher-order QCD corrections.

There is another interesting sum-rule check we can get from the QPM, namely the Adler sum rule. We define the sum [11] as

$$S_A = \int_0^1 \frac{F_2^{\nu n} - F_2^{\nu p}}{x} dx \tag{9.60}$$

and we can see from eqns 9.53 and 9.54 that

$$S_A = \int_0^1 [u(x) - \bar{u}(x) - d(x) + \bar{d}(x)] dx \tag{9.61}$$

Hence, from the definition of valence quarks and eqn 9.56, we expect that

$$S_A = \int_0^1 [u_v(x) - d_v(x)] dx \tag{9.62}$$

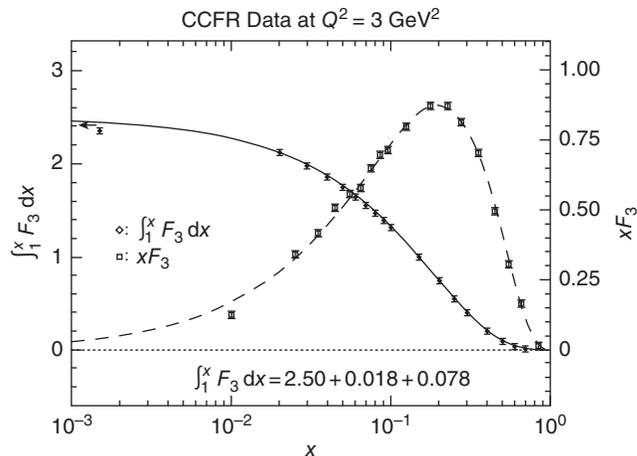


Fig. 9.11 Experimental test of the Gross–Llewellyn Smith sum rule. The dashed line is a fit to the experimental data [98] for $F_3(x)$ and the solid line is the integral.

and since there are two (one) valence u (d) quarks, we expect $S_A = 1$. This prediction is in good agreement with the neutrino DIS data over a range of Q^2 [134], as shown in Fig. 9.12

We will examine the measured shape of the parton distribution function in Section 9.6.6. We can integrate F_2 to get the momentum fraction of the proton carried by all quarks. From eqn 9.55,

$$\begin{aligned} I &= \int_0^1 F_2^{\nu N} dx \\ &= \int_0^1 x[u(x) + d(x) + s(x) + \bar{u}(x) + \bar{d}(x) + \bar{s}(x)] dx \end{aligned} \quad (9.63)$$

Therefore, I is the integral of the momentum-fraction-weighted quark distribution functions; i.e. it represents the total momentum fraction carried by all the quarks. Naively, we might have expected that $I = 1$ for the sum of all the quark flavours, but from the fits of the data [4] shown in Fig. 9.13, we can see that about half of the nucleon momentum is carried by particles that do not feel the electroweak force.¹⁸ In terms of the theory of strong interactions, QCD, we conclude that this missing momentum is carried by gluons. More direct evidence for the existence of gluons will be discussed in Section 9.6.1.

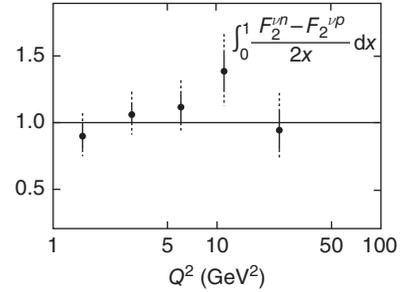
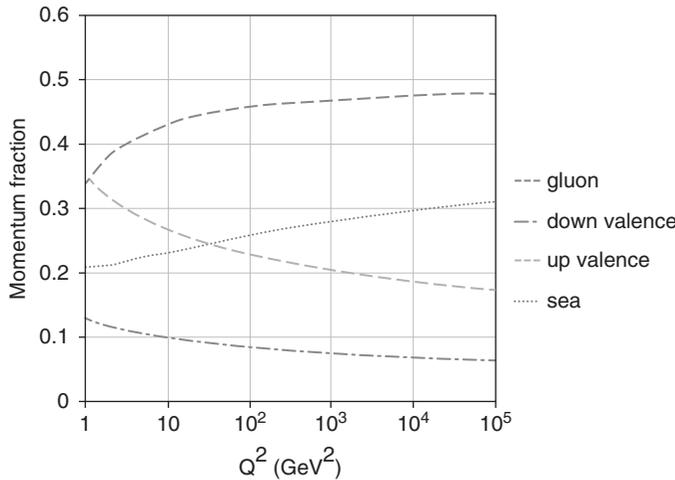


Fig. 9.12 Measured value of S_A (see eqn 9.62) as a function of Q^2 from neutrino DIS [134].

¹⁸Figure 9.13 shows the results for the quarks and gluons. We will discuss how the gluon distribution is determined in Section 9.6.7, but for now the critical point is that the quarks do not carry all the momentum of the proton.

Fig. 9.13 Momentum fraction of the proton carried by different constituents as a function of momentum transfer. The curves are the results of fitting the parton distribution functions [4] and performing the integral in eqn 9.63.

9.5 Charged-lepton probes

Complementary data on the quark distribution functions can be obtained from DIS using charged-lepton (e/μ) beams. As for DIS scattering with neutrino beams, we proceed in the following stages:

- (1) Calculate the cross section for $e\mu$ elastic scattering.
- (2) Generalize to e -quark elastic scattering.
- (3) Generalize to e -nucleon scattering.

9.5.1 Electron–muon elastic scattering

We will calculate the $e\mu$ elastic scattering cross section by analogy with the cross section for $\nu_e e$ scattering. As discussed in Section 9.4.1, we can consider $\nu_e e$ scattering as due to the exchange of a virtual W boson. The strength of the weak coupling constant G_F can be related to the dimensionless coupling constant of the weak interaction, g , and the mass of the W , M_W (see Chapter 7):

$$G_F \approx \frac{g^2}{M_W^2} \quad (9.64)$$

At low values of momentum transfer Q , the strength of the weak interaction is determined by G_F , but at high values of Q , of the order of M_W , we have to allow for the effect of the W propagator, which is to modify the effective strength of the weak interaction to be

$$G_{\text{effective}} \approx \frac{g^2}{M_W^2 + Q^2} \quad (9.65)$$

Now, in the case of $e\mu$ scattering, the interaction is due to the exchange of a photon. For a real photon, $M_\gamma = 0$ and the strength of the electromagnetic interaction is given by the dimensionless fine-structure constant α . Therefore, by analogy with eqn 9.44 for elastic $\nu_e e$ scattering, the cross section for elastic $e\mu$ scattering is given by

$$\frac{d\sigma}{dy} = \frac{\alpha^2 s}{Q^4} F \quad (9.66)$$

where F is a spin factor that we will now evaluate. The spin factor for $\bar{\nu}_e e$ was easy to evaluate since the V–A interaction ensured that there was only one possible spin configuration. In general, if we have different spin configurations, then the recipe for calculating the cross section for unpolarized beams in an experiment where one does not measure the final-state spins is as follows:

- (1) Calculate the cross section for each spin configuration (by calculating the amplitude $A(\theta)$ with the rotation matrices as in Section 2.3).
- (2) Sum over final-state spins.
- (3) Average over initial-state spins.

¹⁹For a pure vector interaction as in the case of electromagnetism (or for a pure axial vector), it can be shown that at high energies there is helicity conservation, i.e. the helicity of an outgoing particle is the same as the incoming particle (see Chapter 6).

The possible spin configurations¹⁹ are shown in Fig. 9.14.

For the reactions in Fig. 9.14(a) and (b), the initial state has $J_z = 0$ and hence it has the same spin factor as for $\nu_e e$ scattering (see eqn. 9.42), giving $A(\theta)$ constant. Similarly, the amplitude for the reaction in Fig. 9.14(c) can be evaluated in the same way as for $\bar{\nu}_e e$ scattering; the electron (muon) states have an initial value of $m = -\frac{1}{2}$

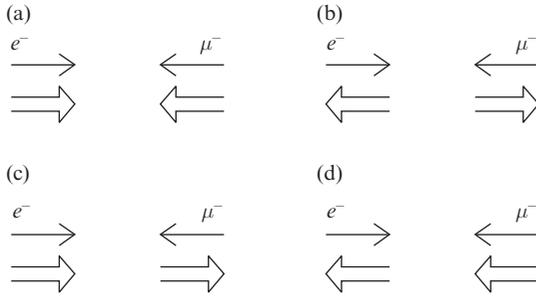


Fig. 9.14 Spin structure for electron–muon scattering.

and a final value of $m' = -\frac{1}{2}$. Therefore, the amplitude is given by the spin- $\frac{1}{2}$ rotation matrix

$$A(\theta) = \left(d_{-1/2, -1/2}^{1/2}\right)^2 = \left(\cos \frac{1}{2}\theta\right)^2 = \frac{1}{2}(1 + \cos \theta) \quad (9.67)$$

and, using eqn 9.32 to change variables, we get

$$A(y) = 1 - y \quad (9.68)$$

Similarly, for the reaction in Fig. 9.14(d), projecting out the spin states gives

$$A(\theta) = \left(d_{1/2, 1/2}^{1/2}\right)^2 = \left(\cos \frac{1}{2}\theta\right)^2 = \frac{1}{2}(1 + \cos \theta) \quad (9.69)$$

and, again using eqn 9.32 to change variables, we get

$$A(y) = 1 - y \quad (9.70)$$

The electromagnetic interaction is not parity-violating and therefore in summing over the final-state spins and averaging over the initial spins, we can assume the same strength of interaction for left- and right-handed helicity states. Therefore, summing the matrix elements squared for the final states and averaging for the initial state, we get the spin factor

$$F = 1 + (1 - y)^2 \quad (9.71)$$

so, from eqn 9.66, the cross section for elastic $e\mu$ scattering is given by

$$\frac{d\sigma}{dy} = \frac{2\pi\alpha^2 s}{Q^4} [1 + (1 - y)^2] \quad (9.72)$$

where the correct numerical factors have to be inserted from a full calculation.²⁰

²⁰See Griffiths in Further Reading.

9.5.2 Electron–quark elastic scattering

Now we can generalize the cross section for elastic $e\mu$ scattering to the case of electron–quark scattering by allowing for the following:

- The fractional electric charge of the quark. We shall denote the charge of the quark of flavour i by q_i .
- If the quark carries a momentum fraction x , then the CMS energy for the e –quark collision is given by the substitution $s \rightarrow \hat{s} = xs$ (see Section 9.4.4).

Therefore, from eqn 9.72, the cross section is given by

$$\frac{d^2\sigma}{dx dy} = \frac{2\pi\alpha^2 xsq_i^2}{Q^4} [1 + (1 - y)^2] \quad (9.73)$$

9.5.3 Electron–nucleon deep inelastic scattering

²¹We consider explicitly e –nucleon scattering here, but of course the analysis is identical for μ –nucleon scattering.

²²This is assuming parity-conserving photon exchange only.

Now we can generalize further to e –nucleon²¹ scattering in the QPM in a similar way as we did for neutrino beams. We assume that e –nucleon DIS can be calculated by adding incoherently the cross sections for scattering off all quark flavours:²²

$$\frac{d^2\sigma}{dx dy} = \frac{2\pi\alpha^2 s}{Q^4} [1 + (1 - y)^2] \sum_i q_i^2 x f_i(x) \quad (9.74)$$

where, as above q_i is the charge of quark flavour i , $f_i(x)$ gives the probability distribution for quark flavour i , and, as usual, x is the momentum fraction of the nucleon carried by the quark. It is convenient to rearrange this formula to give

$$\frac{d^2\sigma}{dx dy} = \frac{4\pi\alpha^2 s}{Q^4} \left[(1 - y) + \frac{1}{2}y^2 \right] \sum_i q_i^2 x f_i(x) \quad (9.75)$$

We now compare this prediction of the QPM with a general phenomenological formula

$$\frac{d^2\sigma}{dx dy} = \frac{4\pi\alpha^2 s}{Q^4} \left[(1 - y)F_2(x, Q^2) + \frac{1}{2}y^2 2xF_1(x, Q^2) \right] \quad (9.76)$$

where F_1 and F_2 (the structure functions) are unknown functions of x and Q^2 . Equating coefficients of $(1 - y)$ and y^2 between eqns 9.75 and 9.76, we have the result [59] that

$$F_2(x, Q^2) = 2xF_1(x, Q^2) \quad (9.77)$$

This important prediction—called the Callan-Gross relation—of the QPM arises because the quarks have spin $\frac{1}{2}$ and interact via a vector (parity-conserving) interaction. The experimental electron scattering

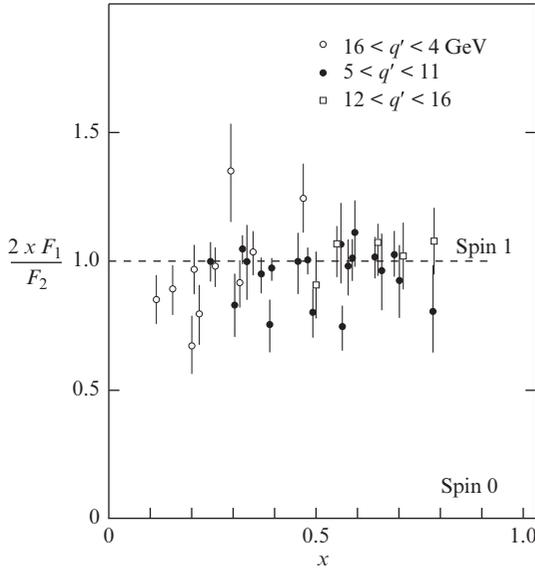


Fig. 9.15 Measured value of the ratio $2xF_1/F_2$ as a function of x . From [117].

data from SLAC, as summarized in [117] and shown in Fig. 9.15, are in good agreement with this prediction, which thus provides further evidence for the existence of spin- $\frac{1}{2}$ quarks.

Another key prediction of the QPM is obtained from a comparison of the QPM prediction (eqn 9.75) with the phenomenological formula (eqn 9.76):

$$F_2(x, Q^2) = \sum_i q_i^2 x f_i(x) \quad (9.78)$$

This means that the QPM predicts that the structure functions depend on x and not on Q^2 , i.e. that they show a scaling behaviour (called Bjorken scaling). This is very directly related to the quarks being point-like particles. If the quarks had a finite size, we would expect the structure functions to be suppressed by a form factor at large Q^2 . The experimental data are shown in Fig. 9.16 and show approximate scaling behaviour.

There is a range of intermediate x values for which F_2 is remarkably constant over a very large variation in Q^2 . This is an extension of the earlier fixed-target data, which gave the first evidence for the QPM. At low values of x , there is a strong rise in F_2 with Q^2 , whereas at large values of x there is a decrease in F_2 with Q^2 . These scaling violations will be discussed qualitatively in the context of the theory of strong interactions, QCD, in Section 9.6.6. It is interesting to note that because $Q^2 = sxy$ at high energy, i.e. large values of s , fixed values of Q^2 and y correspond to lower values of x than at lower energy. This implies that the low- x region is very important for high-energy machines such as the Tevatron and particularly the LHC.

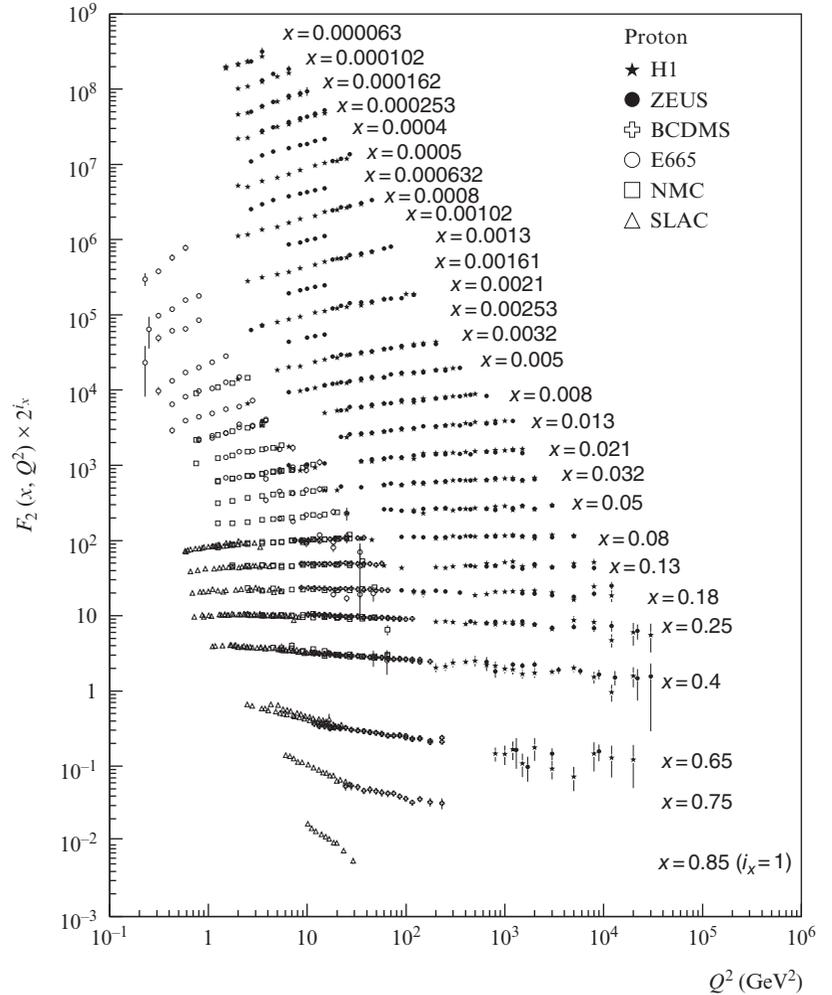


Fig. 9.16 F_2 from fixed-target and HERA scattering experiments From [115].

9.5.4 Further tests of the QPM

In this section, we will look at the QPM prediction for the structure functions and consider some further internal consistency checks of the theory. We start with the QPM prediction for the structure function (eqn 9.78) and we again assume that nucleons contain only the lightest three quark flavours:

$$F_2^{ep}(x, Q^2) = x \left\{ \frac{4}{9}[u(x) + \bar{u}(x)] + \frac{1}{9}[d(x) + \bar{d}(x) + s(x) + \bar{s}(x)] \right\} \quad (9.79)$$

If we assume isospin symmetry as we did when considering neutrino DIS, then we expect $u_p(x) = d_n(x)$ and $\bar{u}(x) = \bar{d}(x)$. Therefore, the equivalent structure function for a neutron target should be

$$F_2^{en}(x, Q^2) = x \left\{ \frac{1}{9}[u(x) + \bar{u}(x) + s(x) + \bar{s}(x)] + \frac{4}{9}[d(x) + \bar{d}(x)] \right\} \quad (9.80)$$

Then, for an isoscalar target with equal numbers of neutrons and protons,

$$F_2^{eN}(x, Q^2) = x \left\{ \frac{5}{18}[u(x) + \bar{u}(x) + d(x) + \bar{d}(x)] + \frac{1}{9}[s(x) + \bar{s}(x)] \right\} \quad (9.81)$$

which we can compare with the equivalent structure function for neutrino scattering (eqn 9.55), and if we ignore the strange quarks (these turn out to be negligible except at very low values of x), then we get the prediction that

$$F_2^{eN}(x, Q^2) = \frac{5}{18} F_2^{\nu N}(x, Q^2) \quad (9.82)$$

where the numerical factor of $\frac{5}{18}$ is just the average value of the square of the quark charges. The experimental data (Fig. 9.17) are in good agreement with the prediction and hence confirm the charge assignments for the u and d quarks.

Another interesting consistency check is given by a comparison of the structure functions for ep and en scattering (eqns 9.79 and 9.80):

$$F_2^{ep}(x, Q^2) - F_2^{en}(x, Q^2) = \frac{1}{3}x[u(x) - d(x) + \bar{u}(x) - \bar{d}(x)] \quad (9.83)$$

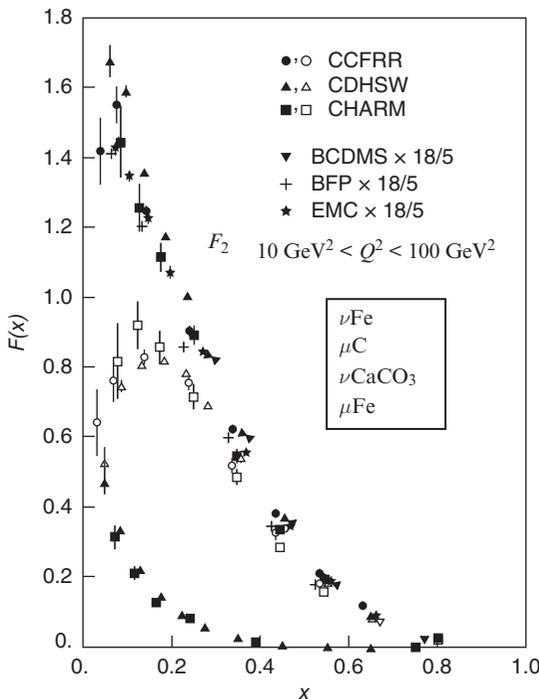


Fig. 9.17 Comparison of $F_2(x)$ from DIS muon and neutrino data. The muon data have been multiplied by a factor of $\frac{18}{5}$ so that the comparison with the neutrino data provides a test [115] of the prediction of eqn 9.81.

We now split the quark distribution functions into valence and sea components:

$$F_2^{ep} - F_2^{en}(x, Q^2) = \frac{1}{3}x\{u_v(x) - d_v(x) + 2[\bar{u}(x) - \bar{d}(x)]\} \quad (9.84)$$

Then, if we integrate eqn 9.84, we obtain the Gottfried sum rule

$$\begin{aligned} I_G &= \int_0^1 \frac{F_2^{ep}(x, Q^2) - F_2^{en}(x, Q^2)}{x} dx \\ &= \int_0^1 \frac{1}{3}\{u_v(x) - d_v(x) + 2[\bar{u}(x) - \bar{d}(x)]\} dx \end{aligned} \quad (9.85)$$

In the QPM, there are two valence up quarks and one valence down quark in the proton, so if we have isospin symmetry, then we should expect $I_G = \frac{1}{3}$. The experimental data [18] are shown in Fig. 9.18 and clearly do not agree with this prediction. This can be understood in terms of the Pauli exclusion principle. Antiquarks are created by the strong interaction at the same time as quarks (the strong interaction conserves quark flavour). The Pauli exclusion principle forbids the creation of a quark in the same quantum state as one of the existing valence quarks. Since there are two (one) valence up (down) quarks in a proton (neutron), it is therefore easier to create $d\bar{d}$ quark pairs than $\bar{u}u$ pairs. Hence the value of I_G should be less than $\frac{1}{3}$.

9.5.5 Electroweak unification at HERA

²³With 27.5 GeV electrons on 920 GeV protons, giving a maximum CMS energy of 318 GeV.

The high CMS energy of the HERA electron–proton collider²³ allowed a vivid demonstration of electroweak unification. Figure 9.19 shows the inclusive cross sections as functions of Q^2 for the processes $e^\pm p \rightarrow e^\pm X$,

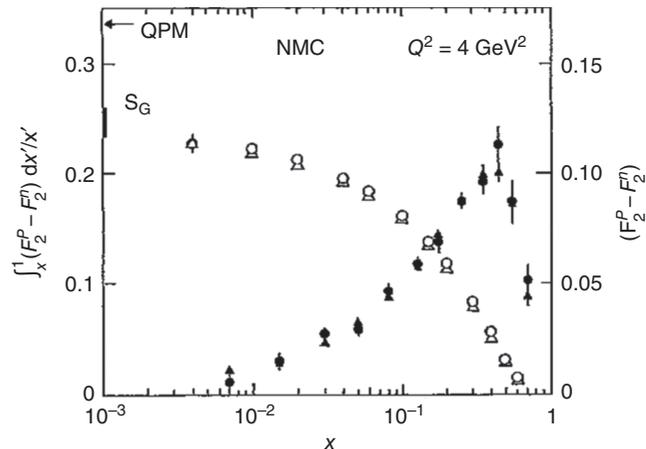


Fig. 9.18 NMC data [18] for $F_2^p - F_2^n$ and the Gottfried sum rule integral (eqn 9.85).

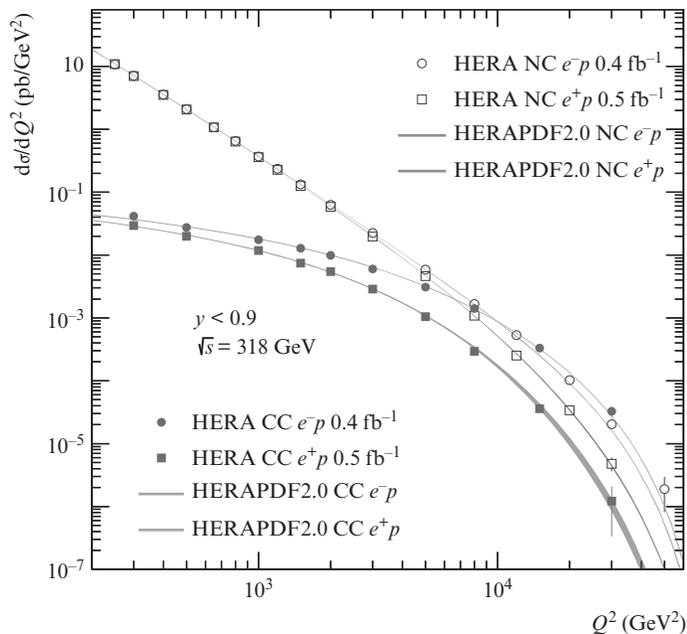


Fig. 9.19 Total cross sections as functions of Q^2 for the processes $ep \rightarrow eX$ (neutral current, CC) and $ep \rightarrow \nu X$ (charged current, CC) measured at the HERA collider [82]. The data from the H1 and ZEUS experiments have been combined. The fit is from the HERAPDF2.0.

mediated by photon and Z^0 exchange (neutral current), and $e^\pm p \rightarrow \nu X$, mediated by W^\pm exchange (charged current). It can be seen that for $Q^2 > M_W^2$, the four cross sections are of comparable magnitude. At smaller Q^2 , the $1/Q^2$ of the photon propagator dominates and the neutral-current cross sections increase rapidly, while the charged-current cross sections become approximately constant in Q^2 .

9.6 QCD introduction

We have seen that about 50% of the momentum in a proton is carried by particles that do not feel the electromagnetic or weak force, and we have tentatively ascribed this momentum to gluons. We will examine direct evidence for gluons in Section 9.6.1, including a measurement of the spin of the gluon. In Section 9.6.2, we will see how to use e^+e^- annihilation data to measure the number of colours carried by quarks. Using this experimental knowledge as input, we can construct a theory of strong interactions called quantum chromodynamics (QCD) based on the symmetry group $SU(3)$. We will give an introduction to this theory in Section 9.6.3. We will then introduce the concept of ‘running coupling constants’ in Section 9.6.4. We will examine this quantitatively for QED and then in a qualitative way for QCD. This leads to the concept of ‘asymptotic freedom’, which allows us to perform perturbation-theory calculations in QCD if we have large values of Q^2 . We then consider some experimental techniques for measuring the strong coupling constant $\alpha_s(Q^2)$ and show that the experimental

measurements are consistent with QCD and provide clear evidence of the running of $\alpha_s(Q^2)$. Finally, we will review some evidence that is sensitive to the choice of the SU(3) group for the colour symmetry.

9.6.1 Direct evidence for gluons

We have seen strong but indirect evidence for the existence of gluons from the momentum sum rule (Section 9.5.4), but there is much more direct evidence from e^+e^- annihilation. In the naive QPM we would only expect 2-jet events, but in QCD we can have multijet events. For example, 3-jet events will be produced by the gluon bremsstrahlung process $e^+e^- \rightarrow q\bar{q}g$. Such events were observed at the PETRA collider at DESY. An example of a clear 3-jet event in the TASSO detector [136] is shown in Fig. 9.20. Many statistical tests were performed to show that these events were not just due to fluctuations of ‘2-jet’ events.

For example, if the events were due to gluon bremsstrahlung, we would expect planar events. The measured transverse momentum out of the plane of the 3 jets was much smaller than that in the plane, confirming

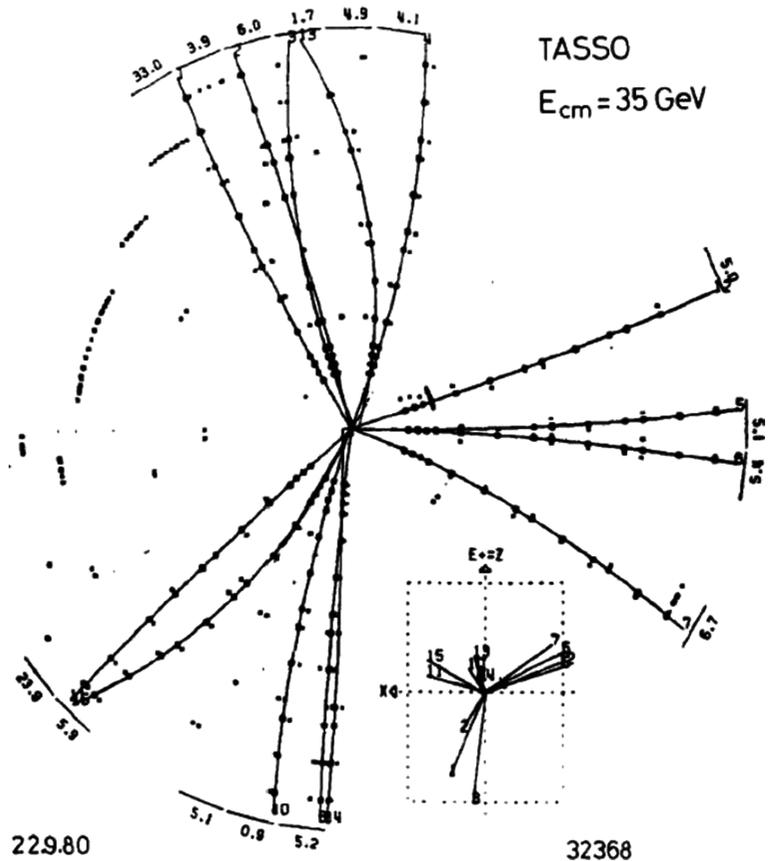


Fig. 9.20 Event display for a 3-jet event from the TASSO experiment at PETRA [136].

the existence of 3-jet events [45]. Apart from confirming the existence of gluons, measurements of the angular distribution of the third jet are sensitive to the gluon spin. In a sample of 3-jet events from the TASSO experiment, the jet energies are ordered $E^1 > E^2 > E^3$; the event is boosted to the CMS frame of jets 2 + 3; the angle $\tilde{\theta}$ is defined as shown in Fig. 9.21. The results from the TASSO experiment [54] are shown in Fig. 9.22, from which it can be seen that the data are consistent with a spin-1 gluon but clearly exclude a spin-0 gluon.

9.6.2 Number of colours

Now that we have seen compelling evidence for the existence of point-like spin- $\frac{1}{2}$ quarks and spin-1 gluons, we are almost ready to consider the theory of strong interactions, QCD. However, we first need to review the evidence that there are three ‘colour’ degrees of freedom for quarks. The classic experimental test of the number of colours is the ratio

$$R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} \quad (9.86)$$

In the QPM, the production of hadrons in e^+e^- interactions proceeds via $q\bar{q}$ states that fragment with unit probability to two jets. Therefore, the fundamental Feynman diagrams for the two processes in eqn 9.86 are the same and the only differences are the charges of the quarks (q_i) and the number of quark colours (N_c). Therefore, in the QPM, we expect

$$R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = N_c \sum_i q_i^2 \quad (9.87)$$

where the sum runs over all available quark flavours (i.e. those for which the CMS energy $E > 2m_i$, with m_i being the quark mass for flavour i). The experimental data [115] are shown in Fig. 9.23.

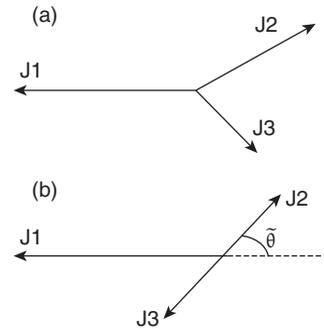


Fig. 9.21 Definition of the Ellis–Karliner angle $\tilde{\theta}$ in 3-jet events. The jets are ordered in decreasing energy and are shown in (a) their common CMS and (b) the CMS of jets 2 and 3.

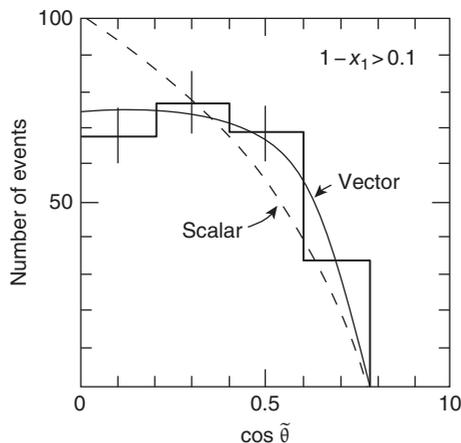


Fig. 9.22 Measurement of the angular distribution in 3-jet events [54] (see the text for the definition of the angular variable used) and comparisons with spin-1 (vector) and spin-0 (scalar) gluons.

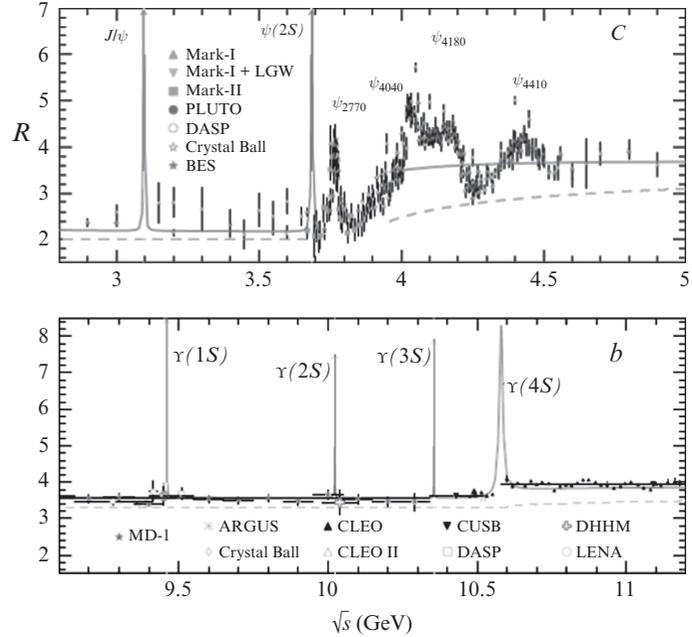


Fig. 9.23 Measurement of the ratio R [115] (see text) and comparisons with the QPM and QCD calculations.

The data show sharp peaks corresponding to hadron resonances. In this region, the quarks are strongly bound and the QPM is not an appropriate approximation. However, away from these resonances, R is approximately constant and shows the step increases as the CMS energy crosses the thresholds for the different quark flavours. The data are clearly inconsistent with $N_c = 1$ and are approximately consistent with $N_c = 3$. We will consider the small discrepancies when we consider QCD theory in the next section.

9.6.3 QCD

Now we have considered the evidence that quarks come in 3 colours and that they interact via the exchange of massless coloured spin-1 gluons, we are ready to consider the theory of strong interactions, QCD. The theory is based on the gauge group $SU(3)$.²⁴ This $SU(3)$ group is associated with 3×3 unitary matrices. A general complex 3×3 matrix requires $3 \times 3 \times 2 = 18$ real parameters to specify it. As the matrices are unitary, they must satisfy $U^\dagger U = I$, or

$$\begin{aligned} \sum_j U_{ij}^\dagger U_{jk} &= \delta_{ik} \\ \sum_j U_{ji}^* U_{jk} &= \delta_{ik} \end{aligned} \quad (9.88)$$

For the diagonal terms, this sum is over terms like $U_{ji}^* U_{ji}$, which must be real (i.e. no imaginary parts), and therefore this yields 3 constraints for

²⁴We have already looked at this group when we were considering the static quark model for hadrons. In that case, we were using an approximate flavour symmetry between the light quarks u , d , and s . Here we will be considering an exact $SU(3)$ colour symmetry.

SU(3). For the off-diagonal terms, both the real and imaginary parts in eqn 9.88 must be equal to 0. There are 6 off-diagonal elements, but if the (i, j) element of the product is 0, then the (j, i) element will also be 0. Therefore, there are 3 off-diagonal elements to consider, each of which provides 2 constraints. So the total number of constraints is $3 + 2 \times 3 = 9$, which leaves 9 parameters. Now consider the determinant of eqn 9.88:

$$\det(U^\dagger U) = \det(I) = 1 \quad (9.89)$$

However, $U^\dagger = U^{T*}$, which implies that $\det(U^\dagger) = \det(U^T)^* = \det(U)^*$, so, from eqn 9.89, $\det(U) \det(U)^* = 1$, i.e. $\det(U) = e^{i\phi}$ (where ϕ is a real number). The special unitary group SU(3) has the additional constraint that $\det(U) = +1$. This additional constraint means that we need 8 parameters and there are 8 generators for SU(3) and therefore 8 gluons. We can represent the three colour states of the quarks as $|r\rangle$, $|b\rangle$, and $|g\rangle$. The use of the term ‘colour’ can lead to some confusion, since it is just a label for a state and has nothing to do with ordinary colour. In particular, a state $|r\rangle|b\rangle|g\rangle$ would not be colourless. We expect there to be 8 generators for the group and therefore 8 colours of gluons from the arguments given above. Although the physics of colour SU(3) is unrelated to that of flavour SU(3), the group theory is of course identical. We can therefore use the results from Chapter 5 to write down the colour wavefunctions of the 8 gluons as shown in Table 9.1. Mathematically, there can be a colour-singlet state

$$\sqrt{\frac{1}{3}} (|r\rangle|\bar{r}\rangle + |b\rangle|\bar{b}\rangle + |g\rangle|\bar{g}\rangle)$$

but if such a gluon state existed in nature, then, as a colour singlet, it could mediate a long-range strong force. Clearly then, there can be no such state for a massless gluon.²⁵

QCD is a similar gauge theory to QED in many ways. We can determine the interactions between quarks and gluons by starting with a free field theory for the quarks and then impose local SU(3) gauge symmetry. Consider²⁶ a local gauge transformation specified by $\Lambda_a(x)$ such that the transformed quark field is

$$\psi'(x) = \exp[i g_s \Lambda_a(x) T_a] \psi(x) \quad (9.90)$$

where g_s will turn out to be the strong-interaction coupling constant, T_a are the generators of SU(3), and an implicit summation over repeated indices is assumed. The infinitesimal transformations of eqn 9.90 are

$$\psi'(x) = [1 + i g_s \Lambda_a(x) T_a] \psi(x) \quad (9.91)$$

In order to keep gauge invariance for the Lagrangian, we are obliged to introduce 8 gauge fields (gluons) $G_\mu^a(x)$. These transform under SU(3) as

$$G_\mu^a(x) \rightarrow G_\mu^a(x) - \partial_\mu \Lambda_a(x) - g_s f_{abc} \Lambda_b(x) G_\mu^c \quad (9.92)$$

$$\begin{aligned} g_1 &= \sqrt{\frac{1}{2}} (|r\rangle|\bar{b}\rangle + |b\rangle|\bar{r}\rangle) \\ g_2 &= -i\sqrt{\frac{1}{2}} (|r\rangle|\bar{b}\rangle - |b\rangle|\bar{r}\rangle) \\ g_3 &= \sqrt{\frac{1}{2}} (|r\rangle|\bar{r}\rangle - |b\rangle|\bar{b}\rangle) \\ g_4 &= \sqrt{\frac{1}{2}} (|b\rangle|\bar{g}\rangle + |g\rangle|\bar{b}\rangle) \\ g_5 &= -i\sqrt{\frac{1}{2}} (|b\rangle|\bar{g}\rangle - |g\rangle|\bar{b}\rangle) \\ g_6 &= \sqrt{\frac{1}{2}} (|g\rangle|\bar{r}\rangle + |r\rangle|\bar{g}\rangle) \\ g_7 &= -i\sqrt{\frac{1}{2}} (|g\rangle|\bar{r}\rangle - |r\rangle|\bar{g}\rangle) \\ g_8 &= \sqrt{\frac{1}{6}} (|r\rangle|\bar{r}\rangle + |b\rangle|\bar{b}\rangle - 2|g\rangle|\bar{g}\rangle) \end{aligned}$$

Table 9.1 Gluon colour wavefunctions.

²⁵A meson state $q\bar{q}$ will be in such a colour-singlet state and can be exchanged between hadrons, but the force is short-range because of the mass of the meson.

²⁶For a more pedagogical look at local gauge transformations and Lagrangians, see Sections 12.1 and 12.4, respectively.

and f_{abc} are the SU(3) structure constants, given by the commutation relations

$$[T_a, T_b] = if_{abc}T_c \tag{9.93}$$

We proceed in a similar way to QED by replacing the partial derivative with the covariant derivative

$$D_\mu = \partial_\mu + ig_s T_a G_\mu^a \tag{9.94}$$

Again following QED, we need to add a kinetic energy term for the gluon fields:

$$L_{\text{gluon}} = F_{\mu\nu}^a F_a^{\mu\nu} \tag{9.95}$$

where the field tensor is given by

$$F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c \tag{9.96}$$

The Lagrangian contains interactions between the quarks and the gluons in a similar way to QED. What is new is that the non-Abelian nature of SU(3) leads to the extra term in eqn 9.96, which, when substituted into the Lagrangian of eqn 9.95, generates terms proportional to G^3 and G^4 . Therefore, the gauge invariance coupled with the fact that SU(3) is non-Abelian implies that there are 3-gluon and 4-gluon vertices. As we shall see in Section 9.6.4, this makes the theory of QCD very different from QED.

We can use the gluon wavefunctions to determine the relative amplitudes for different colour combinations of quark and antiquark scattering:

²⁷Another choice would of course have resulted in the identical amplitude.

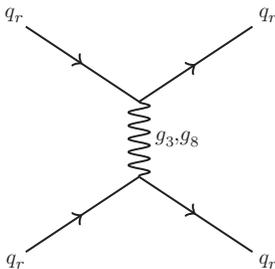


Fig. 9.24 Feynman diagram for quark–quark scattering for red quarks (*t*-channel diagram).

- (1) First let us consider scattering of quarks of identical colour. We make the arbitrary choice of using red quarks²⁷ $q_r q_r \rightarrow q_r q_r$ (see Fig. 9.24). Colour is conserved at each vertex because colour SU(3) is an exact symmetry (in the same way as electric charge is conserved in QED). Therefore, we need to consider gluons that contain $r\bar{r}$ in the colour wavefunction. From Table 9.1, we find g_3 and g_8 , and we can calculate the relative amplitude for g_3 exchange from the normalization constants as $a_3 = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}}$ and that for g_8 exchange as $a_8 = \sqrt{\frac{1}{6}}\sqrt{\frac{1}{6}}$. Adding the two amplitudes gives a factor of $\frac{2}{3}$.
- (2) Similarly, if we consider the case of $q_r q_{\bar{r}} \rightarrow q_r q_{\bar{r}}$, the exchange gluons are g_3 and g_8 as above, but we need to remember the minus sign for antiquarks (in the same way as for negatively charged electrical particles in QED). Therefore, the amplitude is proportional to $-\frac{2}{3}$.
- (3) Next, we will consider the amplitude for scattering of two quarks of different colours (which for convenience we will take to be red and

blue). For the process $rb \rightarrow rb$ (as in Fig. 9.25(a)), we again need to consider the gluons g_3 and g_8 . For g_3 exchange, we need gluons with $r\bar{r}$ and $b\bar{b}$, i.e. g_3 and g_8 , which gives $a_3 = -\sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}} = -\frac{1}{2}$ and $a_8 = \sqrt{\frac{1}{6}}\sqrt{\frac{1}{6}} = \frac{1}{6}$, so the resulting amplitude is $a(rb \rightarrow rb) = -\frac{1}{3}$. The diagram in Fig. 9.25(b) is for the process $rb \rightarrow br$ and involves gluons with $r\bar{b}$ and $b\bar{r}$, i.e. g_1 and g_2 , which gives $a_1 = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}} = \frac{1}{2}$ and $a_2 = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}} = \frac{1}{2}$. Adding the two amplitudes gives $a(rb \rightarrow br) = 1$.

We can now easily write down the amplitude for $q\bar{q}$ states, remembering to use the negative sign for the antiquarks. Thus, $a(r\bar{r} \rightarrow r\bar{r}) = -a(rr \rightarrow rr) = -\frac{2}{3}$ and $a(r\bar{b} \rightarrow r\bar{b}) = -a(rb \rightarrow rb) = \frac{1}{3}$. In a similar way, we can relate $r\bar{r} \rightarrow b\bar{b}$ with $rb \rightarrow br$ because they both involve exchange of gluons g_1 and g_2 . This gives $a(r\bar{r} \rightarrow b\bar{b}) = -a(rb \rightarrow br) = -1$.

The results for the different combinations of colours are summarized in Table 9.2. Now that we have all the pieces, we can easily evaluate the colour factors for $q\bar{q}$ in a colour-singlet configuration. This will turn out to be very interesting, since it gives some insights into the origin of quark confinement. For the colour singlet, we need

$$|q\bar{q}\rangle = \sqrt{\frac{1}{3}}(|r\bar{r}\rangle + |b\bar{b}\rangle + |g\bar{g}\rangle) \quad (9.97)$$

We have to allow for the wavefunction normalization in eqn 9.97 and for the scattering amplitudes given in Table 9.2. Thus, $r\bar{r} \rightarrow r\bar{r}$ gives a factor of $\sqrt{\frac{1}{3}}\sqrt{\frac{1}{3}}(-\frac{2}{3})$. Allowing for the three colours gives a factor of 3, which gives $-\frac{2}{3}$. $r\bar{r} \rightarrow b\bar{b}$ gives $\sqrt{\frac{1}{3}}\sqrt{\frac{1}{3}}(-1)$ and we have to allow for an equivalent term $r\bar{r} \rightarrow g\bar{g}$. Again we allow for the three colours, so the result is $3(-\frac{2}{3}) = -2$. Adding, we obtain the overall result for the colour factor for $q\bar{q}$ in a colour-singlet state, which is²⁸ $-\frac{8}{3}$. We can also calculate colour factors for gluon coupling $g \rightarrow gg$. In the conventional normalization, these colour factors (called ‘Casimir factors’) are given by

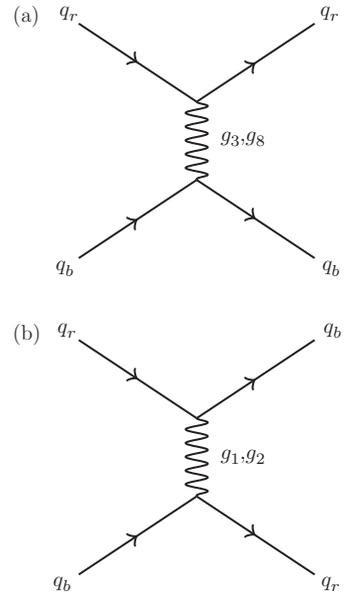


Fig. 9.25 Feynman diagrams for quark–quark scattering for red and blue quarks (t -channel diagrams).

²⁸Different conventions for the definition of the strong-interaction coupling constant can give results for these factors differing by a factor of 2, but the results for any cross section are the same.

Quarks	Antiquarks
$a(rr \rightarrow rr) = \frac{2}{3}$	$a(r\bar{r} \rightarrow r\bar{r}) = -\frac{2}{3}$
$a(rb \rightarrow rb) = -\frac{1}{3}$	$a(r\bar{b} \rightarrow r\bar{b}) = \frac{1}{3}$
$a(rb \rightarrow br) = 1$	$a(r\bar{r} \rightarrow b\bar{b}) = -1$

Table 9.2 Colour factors for quark and antiquark scattering by gluon exchange.

$C_F = \frac{4}{3}$ for $q \rightarrow qq$ and $C_A = 3$ for $g \rightarrow gg$ coupling (see Cooper-Sarkar and Devenish in Further Reading).

Note that we have focused on the contribution to the vertex factor from the gluon colour, but the overall amplitudes will also pick up a factor of the strong-interaction coupling constant and a propagator term. Since the gluons are massless like the photon, the propagator term is $1/q^2$, where q is the 4-momentum transfer. In summary, the overall amplitude for the interaction of the $q\bar{q}$ colour-singlet state is given by

$$a(q^2) \sim -\frac{8 g_s}{3 q^2} \tag{9.98}$$

²⁹When we calculated the Rutherford scattering cross section in Section 9.1, we used the Fourier transform to calculate the amplitude as a function of momentum transfer from the known Coulomb potential. Here we are performing the inverse Fourier transform to determine the potential, starting from the amplitude as a function of momentum transfer.

In the non-relativistic limit, we can Fourier transform the amplitude given in eqn 9.98 from momentum space to position space.²⁹ This gives a potential that is negative and scales with distance as $1/r$. It turns out that this is the only combination of two quarks/antiquarks that gives a negative potential. This gives some explanation of why we find bound $q\bar{q}$ states but not states with net colour like qq . This is suggestive of colour confinement, which states that the only stable hadrons are colour singlets. However, as we will see in the next section, the strong-interaction coupling ‘constant’ becomes large at low values of q^2 , which means that the perturbation theory we have used will no longer be valid, so this result should be considered as a qualitative indication of colour confinement, rather than a proof.

9.6.4 Running coupling constants

In QCD, the effect of the running coupling constant is very important. Even in QED, the fine-structure ‘constant’ is not constant but varies with the scale Q^2 of the reaction being studied. This is due to the effects of shielding by virtual e^+e^- pairs, which reduces the effective strength of the interaction between two charges as the distance increases or, equivalently, as the scale Q^2 decreases. A first naive attempt to calculate the effects of these higher-order corrections results in meaningless infinities. We will see how renormalization theory allows us to overcome this problem.³⁰ Consider the case of $ee \rightarrow ee$. The lowest-order Feynman diagram is shown in Fig. 9.26(a). There is an $O(\alpha)$ correction from the e^+e^- loop diagram shown in Fig. 9.26(b). Evaluation of the loop requires an integration of the 4-momentum running round the loop. It can be shown that the effect of this integral modifies the amplitude by a factor $1 - I(q^2)$, where q is the 4-momentum of the photon and

$$I(q^2) = \frac{\alpha}{3\pi} \int_{m^2}^{\infty} \frac{dp^2}{p^2} - 2\frac{\alpha}{\pi} \int_0^1 dx (1-x) \ln \left[1 - \frac{q^2 x(1-x)}{m^2} \right] \tag{9.99}$$

The result is formally logarithmically divergent and for now we introduce an arbitrary upper limit Λ , but we will see that our final result is

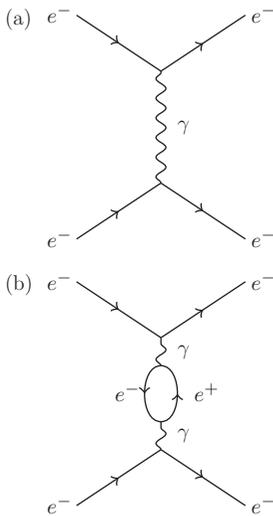


Fig. 9.26 Lowest-order Feynman diagram for $e^-e^- \rightarrow e^-e^-$ scattering (a) and an $O(\alpha)$ correction with an e^+e^- loop (b).

³⁰Our explanation is based on that given in Burcham and Jobes (see Further Reading).

independent of the value of Λ . For the case $-q^2 \gg m^2$, we can evaluate the integral (see Exercise 9.3) as

$$I(q^2) = \frac{\alpha}{3\pi} \ln\left(\frac{\Lambda^2}{-q^2}\right) \quad (9.100)$$

There are also higher-order diagrams to consider, with two, three, four, etc. e^+e^- loops. The effect on the matrix element is to introduce a multiplicative correction factor

$$F = 1 - I(q^2) + [I(q^2)]^2 - [I(q^2)]^3 + \dots \quad (9.101)$$

We can use eqn 9.101 and sum this infinite geometric series to get a factor

$$F = \frac{1}{1 + (\alpha/3\pi) \ln(\Lambda^2/Q^2)} \quad (9.102)$$

where $Q^2 = -q^2$. We can interpret this result by regarding α as the bare coupling constant (which we call α_0), to which the measured coupling constant is related via

$$\alpha = \frac{\alpha_0}{1 + (\alpha/3\pi) \ln(\Lambda^2/Q^2)} \quad (9.103)$$

This result shows that the effective coupling constant depends on the scale Q^2 as

$$\alpha(Q^2) = \frac{\alpha_0}{1 + [\alpha(Q^2)/3\pi] \ln(\Lambda^2/Q^2)} \quad (9.104)$$

We can of course write down an equivalent formula at a different scale μ . We can then combine these two expressions (see Exercise 9.4) to find a relationship between the coupling constants at the two different scales:³¹

$$\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 - [\alpha(\mu^2)/3\pi] \ln(Q^2/\mu^2)} \quad (9.105)$$

This result is remarkable because the dependence on the arbitrary cut-off parameter Λ has disappeared. This means that QED is a fully consistent theory up to any arbitrary energy scale.³² The result of renormalization predicts that the value of $\alpha(Q^2)$ increases slowly with Q^2 . This effect amounts to about 7% when comparing $Q = M_Z$ with $Q = m_e$, and this prediction has been verified. We can understand this result in a qualitative way by noting that the charge of one electron seen by the other will be decreased as a result of screening by the e^+e^- pairs. At higher values of Q^2 , the photon has a shorter wavelength and so penetrates more of the screening charges and ‘sees’ a larger effective charge.

In the case of QCD, we can have analogous shielding effects from the quarks, but there is also an anti-shielding effect from the gluons, since

³¹Here we have only included the effects of e^+e^- loops, but if Q^2 is sufficiently large, we need to consider the effects of other leptons and quarks. There is an additional subtlety in that we have ignored other Feynman diagrams, but it turns out that these cancel exactly.

³²From eqn 9.105, the value of the coupling constant will become infinite at $Q \sim 10^{280}$ MeV, but this would only mean that perturbation theory breaks down, not that the theory is fundamentally wrong.

they have a colour charge (unlike photons, which are neutral). The net result is that the running coupling constant for QCD is given by

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + [\alpha_s(\mu^2)/12\pi](33 - 2n_f) \ln(Q^2/\mu^2)} \tag{9.106}$$

where n_f is the number of active flavours, which depends on the scale Q compared with the mass of the different flavours of quarks. At a particular value of $Q^2 = \Lambda_{\text{QCD}}^2$, the denominator in eqn 9.106 will become equal to 0. This happens when

$$\ln(\Lambda_{\text{QCD}}^2/\mu^2) = -\frac{12\pi}{(33 - 2n_f)\alpha_s(\mu^2)} \tag{9.107}$$

We can then invert eqn 9.107 to obtain an expression for the strong interaction running coupling constant in terms of one unknown parameter Λ_{QCD} :

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2n_f) \ln(Q^2/\Lambda_{\text{QCD}}^2)} \tag{9.108}$$

For $n_f < 17$, we can see from eqn 9.108 that the value of $\alpha_s(Q^2)$ decreases with increasing Q^2 , which already gives an explanation for the success of the naive QPM—at large Q^2 , the value of α_s becomes small enough for perturbation theory to be valid and the QPM can be seen to correspond to the processes that are of lowest order in α_s . At small Q^2 , the value of α_s will approach unity, perturbation theory will break down, and the parton model will not even be a useful approximation.

The QCD-improved parton model can be used to calculate the lowest-order QCD corrections to the naive QPM, so, if sufficiently precise experiments can be performed, the theory can be tested.

The value of $\alpha_s(Q^2)$ can be determined³³ from many different processes, including the following:

- (1) The ratio $R_3 = \sigma(e^+e^- \rightarrow 3 \text{ jets})/\sigma(e^+e^- \rightarrow 2 \text{ jets})$. The numerator involves diagrams (see Fig. 9.27) with gluon emissions from the outgoing quark (antiquark) and the result is therefore proportional to α_s . Therefore, measurements of R_3 at different energies can be used to determine α_s over a broad range of Q^2 .
- (2) $\text{BR}(\tau \rightarrow l\nu_\tau\bar{\nu}_l)$, where l can be either e or μ . To lowest order in α_s , we can use the universality of the charged-current weak interactions. Apart from the two lepton flavours, the τ can decay to $q\bar{q}$. The mass of the τ is too low to allow decays to $c\bar{s}$, and therefore the only hadronic decays (for τ^+) are to $u\bar{d}_C$, where d_C is the Cabibbo-rotated d -quark state. Allowing for 3 quark colours, we get $\text{BR}(\tau \rightarrow l\nu_\tau\bar{\nu}_l) = \frac{1}{5}$. There are similar QCD corrections to the hadronic decay as in the case of e^+e^- annihilation, and therefore precise measurements of the τ branching ratio can be used to determine a value of α_s at a scale $Q^2 = m_\tau^2$.

³³At lowest order in perturbation theory, it is convenient to determine a value of Λ_{QCD} ; however, when higher-order calculations are considered, it is better to measure a value of $\alpha_s(Q^2)$ at a particular scale. Conventionally, the scale used is M_Z .

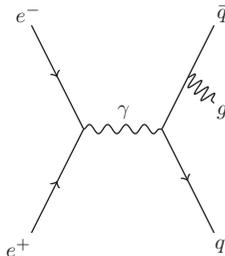


Fig. 9.27 One Feynman diagram for the process $e^+e^- \rightarrow q\bar{q}g$.

- (3) Υ decays. The most common decay mode of the ground state is $\Upsilon \rightarrow 3g$. QED corrections can give the process $\Upsilon \rightarrow \gamma + 2g$, and therefore measurements of the ratio of these two processes depend on the ratio α_s/α .
- (4) Scaling violations in DIS (see Section 9.6.7).

A summary plot [91] of these different determinations of α_s over a large range of Q^2 is shown in Fig. 9.28. The predicted decrease in α_s with increasing Q^2 can clearly be seen. The good agreement between the different measurements and the QCD predictions shows that the theory has now been tested to a precision of better than 1%.

9.6.5 Experimental tests of the gauge structure of QCD

We have seen that there are simple experimental results that demonstrate that quarks come in 3 colours, but there are no equivalent simple demonstrations that gluons come in 8 colours. The question we need to ask is: what is the evidence to justify the choice of the SU(3) gauge group? To some extent, the question has been addressed implicitly by the measurements of $\alpha_s(Q^2)$ as a function of Q^2 , which were consistent with QCD (see Section 9.6.4). However, it is still interesting to see if we can do a more direct test of the choice of the gauge group as SU(3). This can be done in e^+e^- annihilation to 3 or 4 jets. These processes involve the vertices qqg , ggg , and $gggg$, and they are therefore sensitive to the colour factors for qqg and ggg , which in SU(3) are related to the colour factors $C_F = \frac{4}{3}$ and $C_A = 3$ (see Section 9.6.3). The following are some of the variables used in this analysis:

- (1) Charged-particle multiplicity in gluon jets compared with quark jets. As the amplitude for gluon splitting is greater than that for a

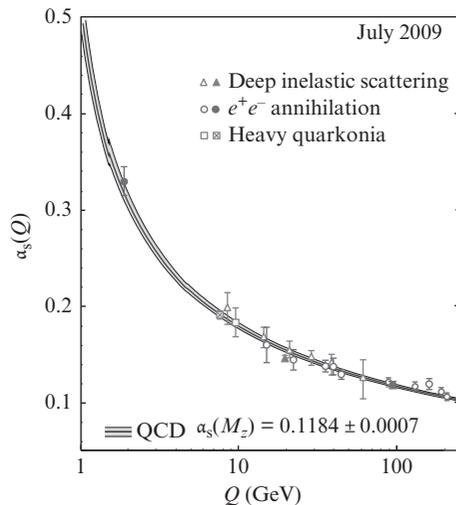


Fig. 9.28 Measurement of α_s with different processes as a function of the scale Q^2 . From [91].

quark to emit a gluon, the multiplicity should be higher in gluon jets than in quark jets. Quark jets from b or c quarks can be tagged by the long lifetimes of the b and c quarks, and the gluon jet will tend to be the lowest-energy jet in 3-jet events because of the bremsstrahlung process.

- (2) The angular distribution in 4-jet events. The angles between the planes of two pairs of jets is sensitive to the qqg , ggg , and $gggg$ vertices and therefore to the colour factors.
- (3) The event shape. This can be parameterized by the variable T , called thrust and defined as

$$T = \max \left(\frac{\sum \mathbf{p}_i \cdot \hat{\mathbf{n}}}{\sum |\mathbf{p}_i|} \right) \tag{9.109}$$

where the sum is over all the calorimeter cells or charged-particle tracks, with momenta p_i . The unit vector $\hat{\mathbf{n}}$ is varied to maximize the value of T . An ideal 2-jet event has $T = 1$ and multijet events will have lower values. The thrust distribution therefore depends on the gluon radiation and is thus sensitive to the colour factors for the different vertices.

The measurements were performed during LEP experiments, with data being taken at (or close to) the Z peak, which provided the highest statistics. The clean and well-defined environment in e^+e^- annihilation and the high energy made it possible to identify clean 3- and 4-jet events and hence to have small systematic errors. The results of a global fit to this data [91] are shown in Fig. 9.29. The data are consistent with the choice of SU(3) for the gauge symmetry and clearly exclude other choices.

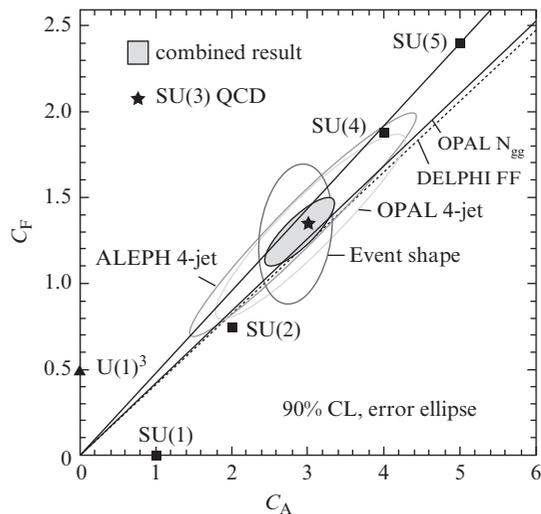


Fig. 9.29 Results of a global fit to e^+e^- data [115] for the colour factors C_F and C_A . The contours from fits to individual analyses are shown and the shaded area is the result for the global fit to all variables. The star represents the QCD prediction.

9.6.6 Experimental fits to the quark distribution functions

From all the DIS data with electron, muon, and neutrino beams, we can perform a fit to determine the quark and antiquark distribution functions. Note that although the most precise data come mainly from electron and muon scattering, the neutrino data are essential to separate antiquarks from quarks. The results of one of these global fits are shown in Fig. 9.30.

We can see that the valence quarks dominate at large x but that the sea quarks are important at very low x . The gluon distribution is also very important at low x , as discussed in the next section.

9.6.7 The gluon distribution function

We have already seen (see Section 9.4.4) that the gluons carry about 50% of the momentum of a nucleon, but the question is how to determine the shape of the gluon distribution function $g(x)$. This cannot be determined as directly as for the quarks, because the gluons carry no electric or weak charge and therefore do not interact directly with photons or W bosons. We can, however, determine the shape of $g(x)$ from the scaling violations, the slow variation of the quark distribution functions with Q^2 (see Fig. 9.16). These scaling violations can be explained in QCD by higher-order corrections to the simple QPM. A quark carries strong charge, so it can emit a virtual gluon, and a gluon also carries strong charge (note the important difference with electromagnetism, where the carrier of the force, the photon, is electrically neutral) and can therefore turn into a

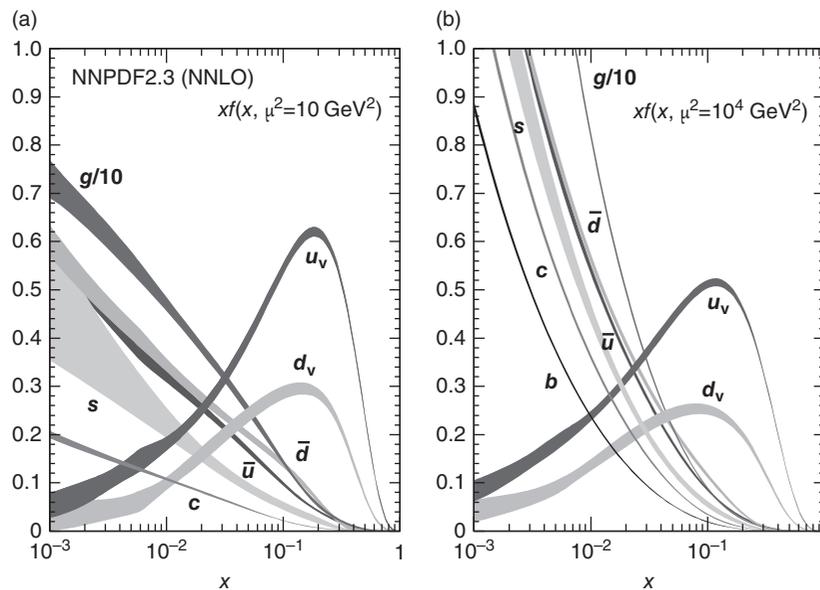


Fig. 9.30 QCD global fits to distribution functions [115] for two different scales: $Q^2 = 10 \text{ GeV}^2$ (a) and $Q^2 = 10^4 \text{ GeV}^2$ (b).

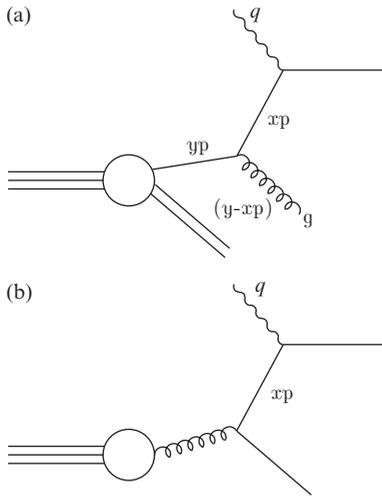


Fig. 9.31 QCD corrections to the QPM.

$q\bar{q}$ pair. Hence we have corrections to the QPM, as illustrated in the Feynman diagrams in Fig. 9.31.

From the process in Fig. 9.31(a), we expect the quarks to move down in momentum, and hence the quark distribution functions will be enhanced at low x and depleted at high x . From the process in Fig. 9.31(b), we expect an enhancement of quarks and antiquarks at low x . How much of this sea of virtual quarks and antiquarks we resolve depends on the wavelength of the probe we use. At longer wavelengths (lower momentum transfer and hence smaller values of Q^2), we do not resolve the quark–antiquark pairs and so they do not give a net contribution. Conversely, at higher values of Q^2 , we have sufficient resolution to resolve them. Hence we expect that as Q^2 increases, $F_2(x, Q^2)$ should increase at low x and decrease at high x . Qualitatively, this is just the behaviour seen in the data (see Fig. 9.16). From a quantitative QCD analysis of this data, we can in fact determine $g(x)$, and the result of such a determination [115] is shown in Fig. 9.30. These scaling violations are proportional to α_s in lowest-order perturbation theory, and therefore the QCD fits can also be used to provide another determination of α_s .

9.7 Hadron–hadron collisions

The naive QPM and the QCD-improved QPM can be easily extended from lepton–hadron collisions to hadron–hadron collisions. Note that these calculations only apply to ‘hard’ processes involving large transverse momentum so that α_s is small enough for perturbation theory to be valid. At low values of Q^2 , the process is too complicated for any QCD predictions to be made and this is a regime in which we are obliged to use simple phenomenological models. For hard processes, the QPM picture for the reaction is shown in Fig. 9.32.

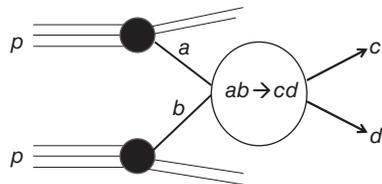


Fig. 9.32 QPM for hadron–hadron collisions. The generic labels $a, b, c,$ and d refer to the type of particles participating in the reaction.

At the parton level, the collision is between a parton with momentum fraction x_a in one hadron and another parton with momentum fraction x_b in the other hadron. The probability of finding a parton at a given momentum fraction x is given by the quark and gluon distribution functions (see Section 9.6.6). The cross section at the parton level can be calculated from perturbation theory (using QED, electroweak theory, or QCD as appropriate). This picture can then be converted into a quantitative prediction in the form of a convolution integral:

$$\sigma(pp \rightarrow cd) = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b f_a(x, Q^2) f_b(x, Q^2) \hat{\sigma}(ab \rightarrow cd) \quad (9.110)$$

where the sum runs over all the parton types, the $f(x, Q^2)$ are the parton distribution functions, $\hat{\sigma}$ refers to the parton–parton collision, which can be calculated, and the integrals are over the parton momentum fractions in the two protons. The formula has been written for the case of pp

collisions, but it can obviously be generalized to other types of hadron–hadron collisions. The momenta of the partons in terms of the CMS energy of the pp collision, \sqrt{s} , are given by $p_a = x_a\sqrt{s}/2$ and $p_b = x_b\sqrt{s}/2$. The square of the CMS energy in the parton–parton collision is therefore given by

$$\hat{s} = E_{\text{total}}^2 - p_{\text{total}}^2 = (x_a + x_b)^2 s/4 - (x_a - x_b)^2 s/4 = x_a x_b s \quad (9.111)$$

9.7.1 Drell–Yan

An application of this formalism is to the Drell–Yan process (lepton–pair production in hadron–hadron collisions). The parton-level process is $q\bar{q} \rightarrow l^+l^-$ and the cross section is given by analogy with $e^+e^- \rightarrow l^+l^-$ as

$$\hat{\sigma}(q_i\bar{q}_i \rightarrow l^+l^-) = \frac{q_i^2}{3} \frac{4\pi\alpha^2}{3\hat{s}} \quad (9.112)$$

where q_i is the charge of the quark in units of e and the extra factor of 3 compared with the equivalent equation for e^+e^- annihilation comes from the fact that only $q\bar{q}$ pairs of the same colour can annihilate to give a virtual photon. By differentiating eqn 9.110 and substituting for $\hat{\sigma}$ from eqn 9.112, we obtain

$$\frac{d^2\sigma(hh \rightarrow l^+l^-)}{dx_a dx_b} = [f_1(x_a)f_2(x_b) + f_1(x_b)f_2(x_a)] \frac{q_i^2}{3} \frac{4\pi\alpha^2}{3\hat{s}} \quad (9.113)$$

It is convenient to change variables to

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \quad \tau = \frac{\hat{s}}{s}$$

where y is the rapidity. From a Jacobian transformation, we obtain

$$\frac{d^2\sigma(hh \rightarrow l^+l^-)}{dy d\tau} = [f_1(x_a)f_2(x_b) + f_1(x_b)f_2(x_a)] \frac{q_i^2}{3} \frac{4\pi\alpha^2}{3s} \quad (9.114)$$

Therefore, if we assume approximate scaling for the quark distribution functions, we should get the same scaled cross section for different values of s . This prediction is compared with experimental data [56] for $pp \rightarrow \mu^+\mu^-X$ in Fig. 9.33. There are many successful predictions for other Drell–Yan processes in $\bar{p}p$ interactions at the CERN $S\bar{p}p$ S collider and the Tevatron and in pp interactions at the LHC. The QCD-improved parton model has been used successfully for many hard processes at the LHC. Some of these applications will be considered in Chapter 13.

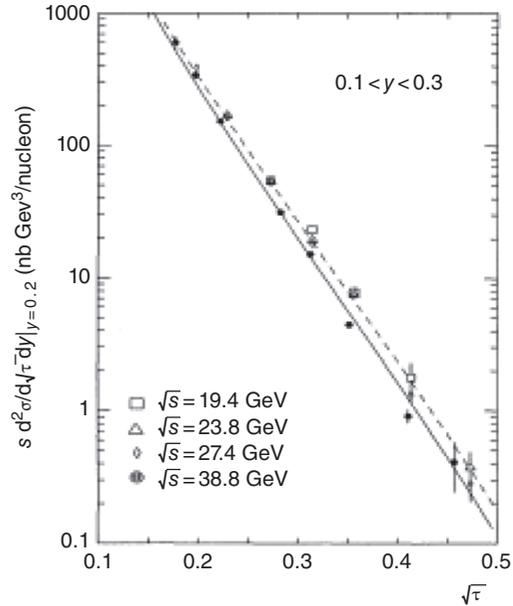


Fig. 9.33 Scaling cross section for $pp \rightarrow \mu^+\mu^-X$ at different values of CMS energy. The data are from the E605 experiment [56].

Chapter summary

- The structure functions show approximate Bjorken scaling behaviour; i.e. F_2 is a function only of x and is independent of Q^2 . This tells us that the scattering occurs off point-like parton constituents (the quarks).
- The Callan–Gross relation (eqn 9.77) tells us that the quarks have spin $\frac{1}{2}$.
- The form of the y distribution in neutrino and antineutrino scattering is consistent with spin- $\frac{1}{2}$ quarks and antiquarks interacting via a parity-violating V–A interaction.
- A comparison of neutrino and electron/muon scattering confirms the quark charge assignments.
- Only about 50% of the momentum of the nucleons is carried by quarks, the rest being carried by gluons.
- From e^+e^- annihilation and hadron–hadron interactions, we have also seen that there are 3 colours of quarks and they interact with spin-1 gluons.
- According to QCD theory, the strong coupling constant $\alpha_s(Q^2)$ decreases with increasing Q^2 . This allows the successful use of perturbative QCD to describe a large number of different experiments, as well as a precise determination of $\alpha_s(M_Z)$.
- All the data are consistent with point-like quarks, but it is always possible that future experiments will reveal that quarks are composite ...
- The parton model can be generalized from lepton–hadron interactions to describe hadron–hadron interactions in ‘hard’ processes, i.e. those in which there is a large momentum transfer (high Q^2). The QCD-improved version of this model will be used to make predictions for hard interactions at LHC energies (see Chapter 13).

Further reading

- Cooper-Sarkar, A. and Devenish, R. (2004). *Deep Inelastic Scattering*. Oxford University Press. This provides a much more detailed and advanced description than that given here.
- Burcham, W. E. and Jobes, M. (1994). *Nuclear and Particle Physics*. Pearson. This includes a more in-depth explanation of renormalization theory.
- Griffiths, D. (2008). *Introduction to Elementary Particles* (2nd revised edn). Wiley-VCH. This gives a very clear introduction to the evaluation of Feynman diagrams.
- Halzen, F. and Martin, A. D. (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley. This is another very good graduate-level textbook explaining the theory of the Standard Model. It contains a clear explanation of how to evaluate Feynman diagrams.
- Söding, P. (2010). On the discovery of the gluon. *EPJ H*, **35**, 3. This is a good summary of this important discovery.

Exercises

- (9.1) Verify the final steps in the derivation of the Rutherford scattering cross section (eqn 9.12).
- (9.2) Consider a model of a quark that has a uniform distribution of electric charge within a radius r_0 . Show that for $qr_0 \ll 1$, the form factor (see eqn 9.16) is given approximately by $F(Q^2) = 1 - aQ^2r_0^2$ and determine the value of the constant a . The HERA DIS data agree with the Standard Model for Q^2 up to values of $\sim 10^4 \text{ GeV}^2$. Use this observation to estimate an upper limit on r_0 .
- (9.3) Perform the integral in eqn 9.99 with the approximation that $-q^2 \gg \Lambda^2$ and verify that you obtain the result given in eqn 9.100.
- (9.4) Use eqn 9.104 to derive eqn 9.105.
Hint: Use $\ln(y/x) = \ln(y/z) + \ln(z/x)$.
- (9.5) Using the data on the y distribution for neutrinos and antineutrinos (see Fig. 9.8), estimate the relative contribution of antiquarks to quarks in the proton at the typical Q^2 scale probed by these reactions. Explain why you would expect this ratio to change with Q^2 .
- (9.6) Use the data shown in Fig. 9.17 to estimate the fraction of the proton momentum carried by gluons. How and why would you expect this fraction to evolve with Q^2 ?
- (9.7) Draw Feynman diagrams to show how charm quarks may be produced in νN and $\bar{\nu} N$ interactions. Explain, with appropriate diagrams and paying particular attention to the signs of the electric charges of the muons, how certain dimuon events may indicate charmed-meson production. Which of νN or $\bar{\nu} N$ is likely to give the larger signal and why?
- (9.8) Consider the Drell–Yan reaction of π^\pm on a carbon target. The carbon nucleus has an equal number of protons and neutrons and thus also an equal number of u and d quarks. Show that in the quark–parton model, the ratio
- $$r = \frac{\sigma(\pi^+ \text{C} \rightarrow \mu^+ \mu^- X)}{\sigma(\pi^- \text{C} \rightarrow \mu^+ \mu^- X)}$$
- equals $\frac{1}{4}$ when \hat{s}/s approaches 1. What value does r have to be for small \hat{s}/s ?
- What are the experimental issues associated with studying Drell–Yan reactions?
- (9.9) Draw Feynman diagrams for the processes $e^+p \rightarrow \bar{\nu}_e X$ and $e^-p \rightarrow \nu_e X$. Hence explain which of these two processes would be expected to have larger cross sections at large Q^2 and x .
- (9.10) Draw a Feynman diagram for charm production in e^-p DIS. Hence explain how measurements of

this process could be used to constrain the gluon parton density function.

- (9.11) The QPM prediction for the ratio R (eqn 9.86) is modified by QCD. According to first-order QCD perturbation theory, the correction is a multiplicative factor $1 + \alpha_s(Q^2)/(3\pi)$. Use this result and the data for R (see Fig. 9.23) to estimate the value of the strong coupling constant $\alpha_s(Q^2)$ for $Q^2 \sim 5 \text{ GeV}^2$.

- (9.12) Consider the Drell–Yan process for protons on a p and deuterium (d) target $pp \rightarrow \mu^+\mu^-X$ and $pd \rightarrow \mu^+\mu^-X$. Explaining any assumptions you make, show that in the QPM, the ratio at given values of x_1 and x_2 is

$$R = \frac{[4u(x_1) + d(x_1)][\bar{u}(x_2) + \bar{d}(x_2)]}{4u(x_1)\bar{u}(x_2) + d(x_1)\bar{d}(x_2)}$$

Oscillations and CP violation in meson systems

10

This chapter and the next deal with the general subject of ‘oscillations’ and CP violation, with this chapter focusing on oscillations and CP violation in meson systems and Chapter 11 dealing with neutrino oscillations. While there are some important differences between oscillations within meson systems and neutrino oscillations, they both demonstrate quantum-mechanical interference over macroscopic distances and allow extremely small mass differences to be measured. Moreover, studies of both the neutral kaon system and neutrino oscillations have produced results challenging the theoretical orthodoxy of their times.

The neutral kaon system has been extensively studied over more than sixty years, mainly in fixed-target experiments. The CP violation phenomenon was discovered in this system in 1962. More recently, B -meson systems have been studied at ‘B-factories’, which have produced spectacular results demonstrating CP -violating effects. There are now four neutral meson systems in which these oscillation and CP -violating processes have been studied: kaons and D , B , and B_s mesons. CP -violating effects have also been observed in decays of charged particles.

CP violation is a necessary condition to understand the observed baryon asymmetry of the universe, but the amount of CP violation observed in the quark sector is too small to explain this effect. However, it is now believed that CP violation in the neutrino sector provides the most plausible explanation, so this subject will be discussed in the context of neutrino oscillations (Chapter 11).

All the CP violation effects observed in the quark sector to date are compatible with the Standard Model. The real interest in CP violation physics in the quark sector is that it usually arises from Feynman diagrams with loops, which naturally makes it very sensitive to new heavy particles coupling within the loop. This means that very precise measurements of CP violation can give access to information about new physics at very high mass scales. This is another illustration of how the indirect search for new physics via precision measurements is complementary to direct searches at machines like the LHC.

10.1 Symmetries	278
10.2 Neutral kaon decays and K_1 and K_2	278
10.3 Mass differences of neutral mesons	281
10.4 Flavour oscillations	282
10.5 CP violation (part 1)	289
10.6 CP violation in the Standard Model	294
10.7 CP violation (part 2)	297
10.8 LHCb measurements	302
Chapter summary	304
Further reading	304
Exercises	304

10.1 Symmetries

Since the role of symmetries is essential to understanding the behaviour of neutral meson systems, the reader should be familiar with these concepts from Chapter 2. The symmetries we discuss here are parity P , charge conjugation C , time reversal T , and their combinations CP and CPT . Parity and charge-conjugation symmetries are conserved for strong and electromagnetic interactions but not for weak interactions. Massless neutrinos and antineutrinos have a definite helicity and only left-handed neutrinos or right-handed antineutrinos participate in the weak interactions: parity is ‘maximally’ violated¹ (see Section 7.2.5).

As will be discussed from Section 10.5 onwards, there is evidence for violation of CP symmetry in the neutral kaon system and in systems containing heavier quarks. Since relativistic field theories predict that the combined symmetry of CPT is conserved, T -violating effects are also expected in these systems.

¹Some of the first evidence for parity violation in weak interactions was provided by the so-called ‘ τ - θ puzzle’, where what is now known as the K^+ was observed to decay into two or three pions, which cannot happen if parity is conserved.

10.2 Neutral kaon decays and K_1 and K_2

The four neutral meson systems that can be studied extensively are listed in Table 10.1. Each of these is the lightest neutral meson containing a particular combination of flavours of quarks, and so the only available decay modes are via the weak interaction. The main thing that makes these particles so fascinating is that there are some decay modes where the same final state is accessible by both the particle and antiparticle, for example $K^0 \rightarrow \pi^+\pi^-$ and $\bar{K}^0 \rightarrow \pi^+\pi^-$. The properties shown in Table 10.1 will be discussed extensively in this chapter—they impart striking differences in the ways in which the four meson systems behave. We start by looking closely at the neutral kaon system, which has two strikingly different lifetimes associated with it, and will then return to this table to discuss the properties of the other meson systems. Because the kaon is light, it has a limited number of decay channels and so is the simplest to consider first.

	ΔM (ps^{-1})	Lifetimes (ps)	
K^0	0.005	89.5	51 000
D^0	0.01	0.41	0.41
B^0	0.504	1.52	1.52
B_s^0	17.7	1.39	1.62

Table 10.1 Neutral meson properties. ΔM is the mass difference between the two mass eigenstates and the lifetimes are given for the two mass eigenstates.

For the initial discussion, we assume that CP is strictly conserved. As just mentioned, the K^0 (quark content $\bar{s}d$) and \bar{K}^0 ($\bar{d}s$) must decay weakly. They decay to two- and three-pion final states, and semileptonically: $K^0 \rightarrow l^+ + \nu_l + \pi^-$ and $\bar{K}^0 \rightarrow l^- + \bar{\nu}_l + \pi^+$, where l is e or μ . The charge of the lepton or pion in these semileptonic decays can be used to determine the strangeness of the decaying kaon. To understand the phenomenology of kaon decays, it is first necessary to examine the properties of the final states under the operation of CP .

The semileptonic final states are CP -conjugate states; in other words, $CP(l^+ + \nu_l + \pi^-) \rightarrow l^- + \bar{\nu}_l + \pi^+$ and vice versa—the two- and three-pion final states are CP eigenstates. The two-pion final states are $\pi^0\pi^0$ and $\pi^+\pi^-$, with a Q value² of ~ 220 MeV. In the $\pi^0\pi^0$ final state, because they are identical bosons, the π^0 s will be in a state of even relative angular momentum, $L = 0, 2$, so $P = [\eta_p(\pi)]^2(-1)^L = +1$

²The Q value is the mass of the parent minus the mass of the decay products and gives the total amount of kinetic energy available in the centre-of-mass frame.

and $C = [\eta_c(\pi^0)]^2 = +1$; hence $CP = +1$. For the $\pi^+\pi^-$ final state, $P = [\eta_p(\pi)]^2(-1)^L = +1$; the operation of C interchanges π^+ and π^- and $C(\pi^+, \pi^-) = (\pi^-, \pi^+)$, so $\eta_c = (-1)^L = +1$ and hence again $CP = +1$.

The Q value of the three-pion final states is only ~ 70 MeV, so this suggests that the most probable value for the relative angular momentum is $L = 0$.³ For the $\pi^0\pi^0\pi^0$ final state, $P = [\eta_p(\pi)]^3 = -1$ and $C = [\eta_c(\pi^0)]^3 = (+1)^3 = +1$; hence $CP = -1$. The $\pi^+\pi^-\pi^0$ final state can be considered as a π^0 combined with the $\pi^+\pi^-$ state. With $CP(\pi^+, \pi^-) = +1$ and $CP(\pi^0) = P(\pi^0)C(\pi^0) = -1$, $CP = -1$ as for the $3\pi^0$ mode.

To summarize, the 2π final states are CP -even and the 3π final states are CP -odd. The important thing to notice is that the Q values of the two- and three-pion decay modes are very different. In particular, the Q value of the three-pion decay mode is only 70 MeV, which leaves very little phase space for these decays, and hence the partial decay rates of K^0 to the two- and three-pion final states will be very different.

K^0 and \bar{K}^0 are not CP eigenstates, although

$$CP|K^0\rangle \rightarrow |\bar{K}^0\rangle, \quad CP|\bar{K}^0\rangle \rightarrow |K^0\rangle$$

Since the K^0 and \bar{K}^0 both decay to the same two- and three-pion final states, it means they are coupled by virtual $|\Delta S| = 2$ second-order weak transitions such as

$$K^0 \leftrightarrow (2\pi) \leftrightarrow \bar{K}^0, \quad K^0 \leftrightarrow (3\pi) \leftrightarrow \bar{K}^0$$

At the quark level, the diagrams that change between K^0 and \bar{K}^0 states are shown in Fig. 10.1; they are often referred to as box diagrams.

This second-order weak coupling ‘mixes’ the K^0 and \bar{K}^0 , meaning that the physical kaon states—the states with definite masses and lifetimes—must evolve as linear superpositions of K^0 and \bar{K}^0 , i.e.

$$|\psi(t)\rangle = a(t)|K^0\rangle + b(t)|\bar{K}^0\rangle$$

The physical states of the neutral kaons that are CP eigenstates can easily be seen to be

$$\begin{aligned} |K_1\rangle &= \sqrt{\frac{1}{2}} (|K^0\rangle + |\bar{K}^0\rangle) \\ |K_2\rangle &= \sqrt{\frac{1}{2}} (|K^0\rangle - |\bar{K}^0\rangle) \end{aligned} \quad (10.1)$$

K_1 and K_2 are orthogonal and expressing these states as linear combinations of K^0 and \bar{K}^0 corresponds to a change of basis from the strong-interaction eigenstates to the weak eigenstates of CP .

³As the momenta of the pions are so low, they would need a very large impact parameter to have $L = 1$.

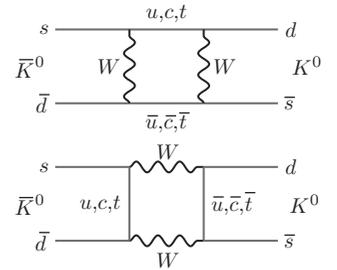


Fig. 10.1 Box diagrams for $\bar{K}^0 \rightarrow K^0$ transitions.

The K_1 and K_2 states contain equal amounts of K^0 and \bar{K}^0 and are eigenstates of even (+1) and odd (-1) CP , respectively, since

$$CP|K_1\rangle \rightarrow \sqrt{\frac{1}{2}}(|\bar{K}^0\rangle + |K^0\rangle) = +|K_1\rangle \quad (10.2)$$

$$CP|K_2\rangle \rightarrow \sqrt{\frac{1}{2}}(|\bar{K}^0\rangle - |K^0\rangle) = -|K_2\rangle \quad (10.3)$$

Since K_1 and K_2 are states of different CP and CP is conserved in their decays, they will decay to different final states with different lifetimes:

$$K_1 \rightarrow 2\pi \quad (CP = +1) \quad (10.4)$$

$$K_2 \rightarrow 3\pi \quad (CP = -1) \quad (10.5)$$

The small phase space available for the $K_2 \rightarrow 3\pi$ decay ($Q \sim 70$ MeV) means that τ_2 , the lifetime of the K_2 , will be much longer than τ_1 , the lifetime of the K_1 .⁴ The K_1 and K_2 are also expected to have different masses. The $K_1 - K_2$ mass difference is very important and is discussed further below; for the present, the discussion will concentrate on the consequences of their different lifetimes.

⁴The heavier D , B , and B_s mesons also mix in a similar way; however, because they are heavier, there are many decay modes with large Q values accessible to each of the CP eigenstates and so the differences in lifetime in the heavy-meson systems are smaller than for kaons.

The strangeness eigenstates K^0 and \bar{K}^0 can be expressed in terms of K_1 and K_2 by inverting eqns 10.1:

$$|K^0\rangle = \sqrt{\frac{1}{2}}(|K_1\rangle + |K_2\rangle) \quad (10.6)$$

$$|\bar{K}^0\rangle = \sqrt{\frac{1}{2}}(|K_1\rangle - |K_2\rangle) \quad (10.7)$$

An initially pure K^0 (or \bar{K}^0) state will be an equal mixture of K_1 and K_2 . If the number of K^0 produced at a (proper) time $t = 0$ is N_0 , then the total number of kaons at a subsequent time t will be

$$N(t) = \frac{1}{2}N_0(e^{-t/\tau_1} + e^{-t/\tau_2}) \quad (10.8)$$

Since $\tau_1 \gg \tau_2$, the K_1 component will die away first and at a time much greater than the lifetime of the K_1 , and only the K_2 component will remain. If a pure K^0 beam is produced, it will be found to contain a rapidly decaying component, K_1 , decaying to two pions, and a slow component, K_2 , decaying to three pions. Although the kaons produced initially are all of the same strangeness, the K_2 component that remains at long times will be an equal mixture of K^0 and \bar{K}^0 . The measured lifetimes of K_1 and K_2 are 89 ps and 51 ns, respectively, corresponding to decay lengths of $c\tau_1 = 2.7$ cm and $c\tau_2 = 15.6$ m. The K_2 lifetime is 600 times the lifetime of the K_1 and at modest energies (i.e. a few GeV) the K_1 component will decay in a few centimetres but the K_2 component will travel many metres before decaying; it is therefore possible to make essentially pure K_2 beams.

10.3 Mass differences of neutral mesons

The discussion in Section 10.2 introduced K_1 and K_2 as the CP eigenstates of the neutral kaon system that arise as a consequence of second-order weak coupling between K^0 and \bar{K}^0 . It was indicated that K_1 and K_2 must have different masses and it is instructive to look more closely at how this arises. We will label the states as $|K\rangle$ etc., but the formalism is valid for any of the neutral meson systems in Table 10.1.

We introduce the formalism in two parts, first by assuming the particles do not decay, then introducing a treatment that allows the mesons to decay. We continue to assume that CP is conserved. Take a meson in a state that is composed of a linear combination of the states $|K^0\rangle$ and $|\bar{K}^0\rangle$ that have definite strangeness (i.e. are eigenfunctions of the strangeness operator). Since $|K^0\rangle$ and $|\bar{K}^0\rangle$ are coupled by second-order weak interactions, we can describe the kaon system by a pair of coupled equations using Schrödinger's time-dependent equation:

$$\begin{aligned} i\frac{\partial}{\partial t} \begin{pmatrix} |K^0\rangle \\ |\bar{K}^0\rangle \end{pmatrix} &= \mathbf{H} \begin{pmatrix} |K^0\rangle \\ |\bar{K}^0\rangle \end{pmatrix} \\ &= \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} \begin{pmatrix} |K^0\rangle \\ |\bar{K}^0\rangle \end{pmatrix} \end{aligned} \quad (10.9)$$

where M is the mass of K^0 and \bar{K}^0 (their masses are identical if we impose CPT invariance). The off-diagonal term $M_{12} = \langle K^0 | H_{\text{weak}} | \bar{K}^0 \rangle$ represents the second-order weak coupling between $|K^0\rangle$ and $|\bar{K}^0\rangle$. The matrix \mathbf{H} represents the Hamiltonian of the system and so is Hermitian. If we instead consider the state as a combination of $|K_1\rangle$ and $|K_2\rangle$ (which are eigenfunctions of the CP operator), the Schrödinger equation can be written in terms of a pair of equations that are no longer coupled:

$$i\frac{\partial}{\partial t} \begin{pmatrix} |K_1\rangle \\ |K_2\rangle \end{pmatrix} = \mathbf{H}' \begin{pmatrix} |K_1\rangle \\ |K_2\rangle \end{pmatrix} = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} |K_1\rangle \\ |K_2\rangle \end{pmatrix} \quad (10.10)$$

The elements of the matrix \mathbf{H}' give the masses of the K_1 and K_2 and can be found from the eigenvalues of \mathbf{H} , which are $M \pm |M_{12}|$. This is how the mass splitting of the K_1 and K_2 arises. The eigenvectors of \mathbf{H} give the states K_1 and K_2 in terms of K^0 and \bar{K}^0 as in eqns 10.1.

We now turn to the case where the particles are able to decay. The time evolution of a neutral meson wavefunction may be written as

$$|K^0(t)\rangle = e^{-iMt} e^{-t/2\tau} |K^0\rangle \quad (10.11)$$

where the first exponential is the usual plane wave for a state with energy $E = M$. The second exponential is imposed⁵ to give the exponential decay for a state with proper lifetime τ (i.e. width $\Gamma = 1/\tau$ in natural units), so that

$$|\langle K^0 | K^0 \rangle|^2 \propto e^{-t/\tau} \quad (= e^{-\Gamma t})$$

⁵Although this is standard practice for considering meson decays, it is a somewhat non-standard use of quantum mechanics. This second exponential means that the state as written does not remain normalized; there is a further piece to the wavefunction, not written here, that represents the part of the state that has decayed.

We can generalize eqn 10.11 to a two-state system:

$$\begin{pmatrix} |K^0(t)\rangle \\ |\bar{K}^0(t)\rangle \end{pmatrix} = \Sigma \begin{pmatrix} |K^0\rangle \\ |\bar{K}^0\rangle \end{pmatrix}, \quad \text{where } \Sigma = e^{-i\mathbf{M}t - \mathbf{\Gamma}t/2} \quad (10.12)$$

\mathbf{M} and $\mathbf{\Gamma}$ are 2×2 matrices encoding the time evolution of the two-state system. $|K^0\rangle$ and its antiparticle $|\bar{K}^0\rangle$ are the *flavour* eigenstates. The off-diagonal terms describe the transitions [mixing] between the meson and antimeson. We can apply Schrödinger's equation, $i d\psi/dt = H\psi$, to identify the Hamiltonian, as we did previously:

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma}$$

Because of the way we have introduced the decaying states, \mathbf{H} is not Hermitian; however, since any matrix \mathbf{A} can be written in the form $\mathbf{A} = \mathbf{H}_1 + i\mathbf{H}_2$, where \mathbf{H}_1 and \mathbf{H}_2 are Hermitian, it follows that \mathbf{M} and $\mathbf{\Gamma}$ are Hermitian matrices. Also, we impose *CPT* invariance (a particle and its antiparticle have identical mass and lifetime) to give

$$\begin{aligned} M_{21} &= M_{12}^*, & \Gamma_{21} &= \Gamma_{12}^* \\ M_{11} &= M_{22} = M, & \Gamma_{11} &= \Gamma_{22} = \Gamma \end{aligned}$$

So the Hamiltonian simplifies to

$$\mathbf{H} = \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix} \quad (10.13)$$

The five fundamental quantities describing the mixing system are M , Γ , $|M_{12}|$, $|\Gamma_{12}|$, and their relative phase $\arg(M_{12}/\Gamma_{12})$. M_{12} and Γ_{12} cannot be fully determined, since there is an arbitrary, unobservable phase in the wavefunction of $|K^0\rangle$. \mathbf{H} can be diagonalized as above to give the masses and lifetimes of the K_1 and K_2 in terms of these quantities. The mass splitting is a more complicated expression than the one we derived above for the non-decaying mesons, but is still approximately $\pm|M_{12}|$.

10.4 Flavour oscillations

Carrying on with the discussion from Section 10.3, we continue to use the kaon system as an example, but these results are applicable to all the meson systems. Since M_{12} arises from second-order weak interactions, the mass difference $\Delta M = M_1 - M_2 \sim 2M_{12}$ is very small. It can be measured by examining the strangeness of a decaying K^0 beam. Consider an initially pure K^0 state at proper time $t = 0$.⁶

⁶Time in the kaon centre-of-mass system.

$$|\psi(t=0)\rangle = |K^0\rangle = \frac{1}{\sqrt{2}}(|K_1\rangle + |K_2\rangle)$$

The state evolves in time as

$$|\psi(t)\rangle = \sqrt{\frac{1}{2}} \left(e^{(-iM_1 - \Gamma_1/2)t} |K_1\rangle + e^{(-iM_2 - \Gamma_2/2)t} |K_2\rangle \right) \quad (10.14)$$

where Γ_1 and Γ_2 are the decay widths of K_1 and K_2 . $|\psi(t)\rangle$ can be re-expressed in terms of $|K^0\rangle$ and $|\bar{K}^0\rangle$ by using eqns 10.1:

$$\begin{aligned} |\psi(t)\rangle &= \frac{1}{2} \left(e^{(-iM_1 - \Gamma_1/2)t} + e^{(-iM_2 - \Gamma_2/2)t} \right) |K^0\rangle \\ &\quad + \frac{1}{2} \left(e^{(-iM_1 - \Gamma_1/2)t} - e^{(-iM_2 - \Gamma_2/2)t} \right) |\bar{K}^0\rangle \end{aligned} \quad (10.15)$$

Interference will occur between the terms of frequencies M_1 and M_2 in the amplitudes of $|K^0\rangle$ and $|\bar{K}^0\rangle$. The K^0 intensity at time t is

$$\begin{aligned} I(K^0) &= |\langle K^0 | \psi(t) \rangle|^2 \\ &= \frac{1}{4} \left(e^{-\Gamma_1 t} + e^{-\Gamma_2 t} \right) + \frac{1}{2} \left(e^{-(\Gamma_1/2)t - (\Gamma_2/2)t} \right) \cos(\Delta M t) \end{aligned} \quad (10.16)$$

Similarly, the intensity of \bar{K}^0 from the same initial K^0 can be obtained from $|\langle \bar{K}^0 | \psi(t) \rangle|^2$:

$$I(\bar{K}^0) = \frac{1}{4} \left(e^{-\Gamma_1 t} + e^{-\Gamma_2 t} \right) - \frac{1}{2} \left(e^{-(\Gamma_1/2)t - (\Gamma_2/2)t} \right) \cos(\Delta M t) \quad (10.17)$$

Equations 10.16 and 10.17 imply that the strangeness content of a beam initially containing strangeness $+1$ K^0 s will oscillate with time (or distance in the laboratory). The oscillations will be observable if $\Delta M \tau_1 \sim 1$ or greater. Figure 10.2(a) illustrates the behaviour of the K^0 and \bar{K}^0 intensities given by eqns 10.16 and 10.17. Oscillations are observable for a few K_1 lifetimes. At times $t \gg \tau_1$, the K_1 component will have entirely decayed away, leaving only K_2 , and the beam will be an equal mixture of K^0 and \bar{K}^0 and zero net strangeness. The total kaon intensity at any time is given by the sum of eqns 10.16 and 10.17 and is as given by eqn 10.8, as it must be.

The features of the oscillations of each of the four systems are remarkably different owing to the different values of the parameters in Table 10.1. The intensities from eqns 10.16 and 10.17 are plotted for each of the meson systems in Fig. 10.2(a–d), and we describe each in the following subsections.

10.4.1 K^0 – \bar{K}^0 oscillations

Strangeness oscillations can be observed experimentally by starting with a pure $S = +1$ K^0 beam produced, for example, by $K^+ + n \rightarrow K^0 + p$ and tagging the strangeness of the decaying kaon by using the semileptonic decays $K^0 \rightarrow l^+ + \nu + \pi^-$ and $\bar{K}^0 \rightarrow l^- + \bar{\nu} + \pi^+$. The $K_1 - K_2$ mass difference can be deduced from the period of the oscillation: $\Delta M = 3.483 \times 10^{-12}$ MeV and $\Delta M \tau_1 = 0.49$. Being less than one part in 10^{14}

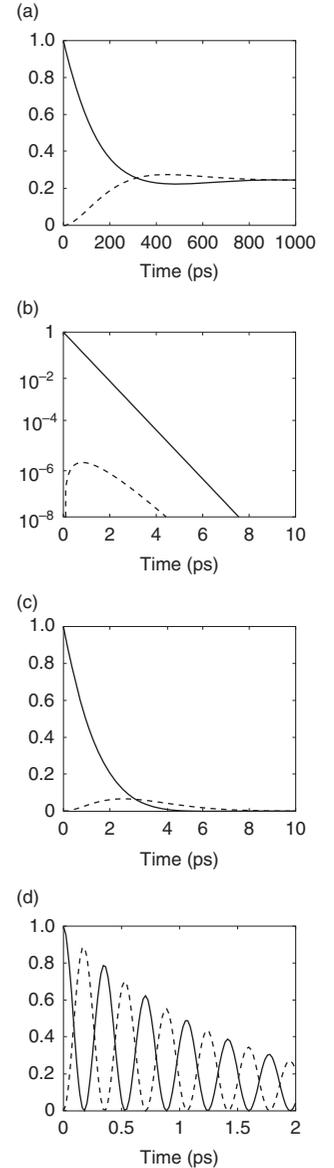


Fig. 10.2 Intensities of mesons (full lines) and antimesons (dashed lines) as functions of proper time after the production of a pure meson state according to eqns 10.16 and 10.17 for the kaon (a), D^0 (b), B^0 (c), and B_s^0 (d) systems.

of the K^0 mass, ΔM is truly tiny, and is the smallest mass difference that has ever been measured.

A rough dimensional estimate of ΔM can be made by considering the ‘box’ diagrams for $\Delta S = 2$ $K^0 \rightarrow \bar{K}^0$ transitions shown in Fig. 10.1. The transition is second-order weak with s - and d -quark to u -, c -, or t -quark transitions occurring at the vertices. The s and d to u and c couplings are the most important at the relatively low energy scale of the kaons, so

$$\Delta M \sim \langle K^0 | H_{\text{weak}} | \bar{K}^0 \rangle \sim G_{\text{F}}^2 m_K^5 \cos^2 \theta_C \sin^2 \theta_C \quad (10.18)$$

⁷This could equally have been written in terms of CKM elements but in two generations, $V_{ud} = V_{cs} = \cos \theta_C$ and $V_{cd} = -V_{us} = -\sin \theta_C$.

where θ_C is the Cabibbo angle.⁷ The use of the kaon mass m_K ensures that eqn 10.18 has the correct dimensions. A full calculation involves the evaluation of loop Feynman diagrams (the box diagrams shown in Fig. 10.1), with the result

$$\Delta M \approx \frac{G_{\text{F}}^2 f_K^2 m_K}{3\pi^2} \cos^2 \theta_C \sin^2 \theta_C \frac{(m_c^2 - m_u^2)^2}{m_c^2} \quad (10.19)$$

⁸A similar argument applies to the rare decay $K_{\text{L}}^0 \rightarrow \mu^+ \mu^-$. Without the presence of the c quark, the branching ratio would be expected to be $O(\alpha^2)$, where α is the fine-structure constant. Allowing for the existence of the c quark, the branching ratio

$$\text{BR} \sim \alpha^2 (m_c^2 - m_u^2) / M_W^2$$

where m_c , m_u , and M_W are the masses of the c quark, u quark, and W boson, respectively.

⁹The first evidence for D -meson oscillations was found at the B-factories and at the Tevatron.

¹⁰We implicitly assume the inclusion of charge-conjugate states.

¹¹In a two-generation approximation, the CKM matrix element is $\cos \theta_C$ and the small value of the Cabibbo angle results in a large value of this element. Hence these decay modes are called ‘Cabibbo-favoured’.

where $f_K \sim 170$ MeV is the experimentally determined kaon decay factor. In a model without charm, the prediction gives a result ~ 4000 times higher than the experimental result, but good agreement is obtained if the charm quark is included in the calculation.⁸ Thus measurement of neutral kaons gave indirect evidence for the existence of the c quark and provided a rough prediction for its mass before it was discovered. This is a classic example of how precision low-energy measurements are sensitive via loop diagrams to physics at much higher mass scales.

10.4.2 D^0 – \bar{D}^0 oscillations

In general, we should expect to observe D^0 – \bar{D}^0 oscillations because the same type of box diagrams responsible for kaon oscillations (Fig. 10.1) will cause $D^0 \leftrightarrow \bar{D}^0$ transitions. However it turns out to be much more difficult to observe oscillations involving charm quarks. First (as for B mesons), there are so many possible decay modes that the lifetimes of the D_1 ‘short-lived’ and D_2 ‘long-lived’ mesons will be very similar (unlike the case for kaons), so we cannot produce a pure sample of the long-lived neutral D mesons. Also, the particular values of the CKM elements cause the rate of oscillations of D mesons to be much slower than their decays, as illustrated in Fig. 10.2(b). Therefore, we need to observe D mesons over a time period of many lifetimes to be able to observe oscillations. This implies that we need very high-statistics samples of D mesons, and the best place to obtain these is using the LHCb experiment at the LHC.⁹ The oscillations were studied by measuring¹⁰ the time dependence of the ratio

$$R = \frac{N(D^0 \rightarrow K^+ \pi^-)}{N(D^0 \rightarrow K^- \pi^+)} \quad (10.20)$$

The decays $D^0 \rightarrow K^+ \pi^-$ involve the quark-level transitions $c \rightarrow s$ and $u \rightarrow d$, both of which have a large CKM mixing angle.¹¹ The decays

$D^0 \rightarrow K^- \pi^+$ involve the quark-level transitions $c \rightarrow d$ and $u \rightarrow s$, both of which depend on the Cabibbo angle as $\sin \theta_C$ and hence these rare decays are called ‘doubly Cabibbo-suppressed’. We can also have decays of D^0 to the ‘wrong-sign’ $K\pi$ decays if the D^0 oscillates into a \bar{D}^0 that then decays by a ‘Cabibbo-allowed’ mode to $K^- \pi^+$. Therefore, the signature for D^0 oscillations will be the ratio of ‘wrong-sign’ to ‘right-sign’ decays given by R (see eqn 10.20) increasing with time as more oscillations occur. From an experimental perspective, we need to

- ‘Tag’ the initial flavour of the D^0 (is it a D^0 or a \bar{D}^0 ?) at production.
- Measure the lifetime of the D^0 .
- Identify the decay products of the D^0 to find $K^\pm \pi^\mp$ events.

The flavour tagging is done by selecting decays $D^{*\pm} \rightarrow D^0 \pi^\pm$, where the charge of the π^\pm determines the flavour of the neutral D^0 at production. This is done by first reconstructing D^0 mesons and then combining them with π^\pm and looking for a narrow peak in the invariant mass spectrum of $\Delta m = m(D^0 \pi^\pm) - m(D^0)$. The lifetime of the D^0 is determined by measuring the flight path L and momentum¹² p , so that the decay time is given by $t = Lm_{D^0}/p$. The measured value [103] of the ratio R is shown in Fig. 10.3. The value of R is increasing with time, as expected for oscillations. The oscillations are so slow compared with the lifetime that only a fraction of an oscillation period can be observed.

¹²The momentum of the neutral particle is reconstructed from the momenta of the charged decay products.

10.4.3 $B^0-\bar{B}^0$ mixing and oscillations

Neutral B mesons come in two varieties, B_d^0 and B_s^0 , containing a \bar{b} and either a d or an s quark, respectively.¹³ The masses of the strong-interaction eigenstates are $M_B = 5.28$ GeV and $M_{B_s} = 5.37$ GeV.

¹³The usual convention is that B^0 means B_d^0 and the subscript d is omitted. B^0 contains a \bar{b} antiquark, whereas D^0 contains a c quark.

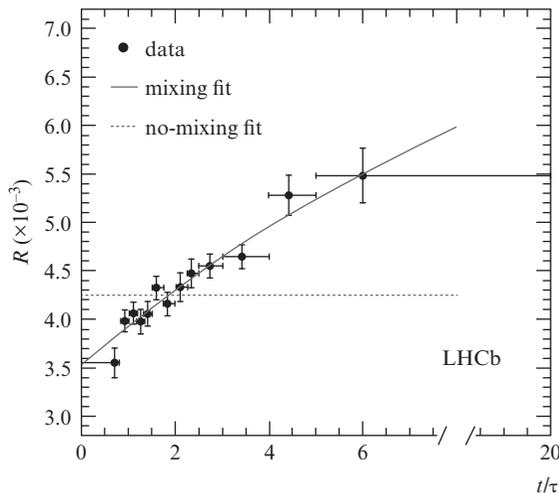


Fig. 10.3 The ratio R of ‘wrong-sign’ to ‘right-sign’ decay modes as a function of decay time divided by the mean lifetime. The curve shows a fit allowing for D^0 oscillations and is in good agreement with the data. The dashed horizontal line shows the ‘no-oscillation’ model. From [103]. The data point at the largest value of t/τ covers the range $5 < t/\tau < 20$.

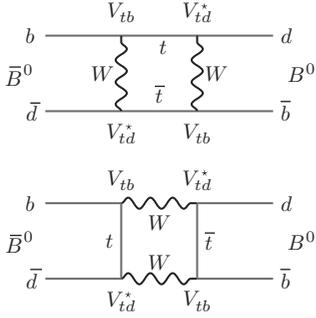


Fig. 10.4 Box diagram for $B_d^0 - \bar{B}_d^0$ transitions.

The large mass means that they have many possible decay modes and short lifetimes, and so (as with the D mesons, but unlike the kaons), the B_1 and B_2 cannot be studied separately.

As discussed in Section 10.3, the condition for observing oscillations is that $\Delta M \tau$ is of the order of unity or greater, where ΔM is the mass difference between the physical eigenstates and τ is the lifetime. Figure 10.4 shows two diagrams for $B_d^0 - \bar{B}_d^0$ transitions. By direct analogy with eqn 10.18, the $B_1 - B_2$ mass difference must be proportional to M_B^5 times the appropriate CKM elements, i.e.

$$\Delta M_B = M_{B_1} - M_{B_2} \approx G_F^2 M_B^5 |V_{tb}|^2 |V_{td}|^2$$

B mesons have a relatively fast oscillation compared with their lifetime, which makes oscillation studies easier than for D mesons, because of two factors: first, only top-type quarks run in the ‘box’ diagram responsible for oscillations and, second, the lifetime is relatively long because of the small value of the CKM matrix elements in the decays (V_{cb} and V_{ub}).

By assuming the lifetimes of B_1 and B_2 are the same, the expressions for the $B^0 \leftrightarrow \bar{B}^0$ transition probabilities are somewhat simpler than those for $K^0 \leftrightarrow \bar{K}^0$ (eqns 10.16 and 10.17):

$$\begin{aligned} P(B^0 \rightarrow B^0) &= \frac{1}{2} e^{-\Gamma_B t} [1 + \cos(\Delta M_B t)] \\ P(B^0 \rightarrow \bar{B}^0) &= \frac{1}{2} e^{-\Gamma_B t} [1 - \cos(\Delta M_B t)] \end{aligned} \quad (10.21)$$

where $\Gamma_B = 1/\tau_B$. Equations 10.21 represent the probabilities of observing a B^0 or \bar{B}^0 at some time t after a B^0 or \bar{B}^0 has been created.

We can tag the flavour of a decaying B meson using the semileptonic decay modes. Therefore, if we produce a $B^0 \bar{B}^0$ pair, we can use events in which both B s decay semileptonically and identify events with same-sign leptons (SS) as having oscillated and those with opposite-sign leptons (OS) as not having oscillated. We define the usual asymmetry $A = (\text{OS} - \text{SS})/(\text{OS} + \text{SS})$, and from eqn 10.21 we can see that $A = \cos(\Delta M_B t)$. Therefore, if we can measure the frequency of these oscillations, we can determine the mass difference between the light and heavy states.

$B^0 - \bar{B}^0$ mixing can be detected by producing $B^0 \bar{B}^0$ pairs, for example in an $e^+ e^-$ collider, and observing their semileptonic decays. At the quark level,

$$\begin{aligned} b &\rightarrow c + W^- \rightarrow c + \mu^- \bar{\nu}_\mu \\ \bar{b} &\rightarrow \bar{c} + W^+ \rightarrow \bar{c} + \mu^+ \nu_\mu \end{aligned}$$

and one signature for $B^0 - \bar{B}^0$ oscillations is the observation of like-sign muon pairs: $\mu^\pm \mu^\pm$. Figure 10.5 shows the fraction of all muon pairs that have the same sign,

$$F = \frac{N^{++} + N^{--}}{N^{++} + N^{--} + N^{+-}}$$

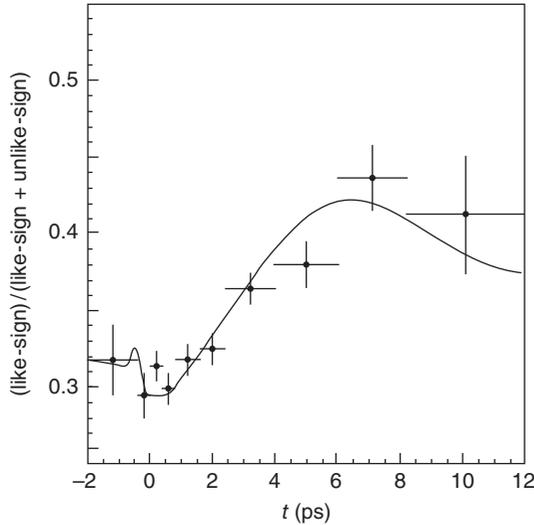


Fig. 10.5 Fraction of like-sign muons due to B^0 - \bar{B}^0 mixing versus proper time in the B^0 system observed by the DELPHI experiment at LEP [72]. The curve is the result of a prediction with $\Delta M_{B_d} = 0.480 \text{ ps}^{-1}$ ($3.159 \times 10^{-10} \text{ MeV}$).

versus proper time in the B^0 system measured by the DELPHI experiment at the LEP e^+e^- collider. There is clearly a significant, time-dependent, excess of like-sign pairs, which is attributable to B^0 - \bar{B}^0 mixing. It can be deduced from these data that $\Delta M_{B_d} \sim 3 \times 10^{-10} \text{ MeV}$ and $\Delta M_{B_d} \tau \sim 0.7$.¹⁴

10.4.4 B_s oscillations

The study of the neutral B_s system has only recently become possible, because the best place to study it is in high-energy hadron colliders such as the LHC (and the Tevatron before that). The LHCb detector at the LHC is a special facility for studying B mesons—Section 10.7.3 describes many of the general features of the LHCb detector.

To study B_s oscillations, we first need to identify the flavour of the neutral B hadrons at their decay to know if they are either B_s^0 or \bar{B}_s^0 . Second, we use the fact that the B hadrons arise from $b\bar{b}$ production to infer the flavour that the signal B hadron had at creation from the flavour of the other B hadron in the event (this is called opposite-side tagging). This can be achieved by several algorithms including the charge of the leptons (e or μ)¹⁵ from semileptonic decays.¹⁶ Same-side tagging (SST) can also be used to identify the flavour of the B_s^0 .¹⁷ The B_s^0 are identified by a flavour-specific decay mode and the D_s^- are identified by decay modes such as $D_s^- \rightarrow \phi\pi^-$ with $\phi \rightarrow K^+K^-$. The very good K/π separation provided by the RICH detectors reduces the combinatorial backgrounds in the mass reconstruction. The pion charge identifies the flavour of the B_s^0 .

An event with a B_s^0 and a \bar{B}_s^0 would be identified as non-mixed, whereas an event with two B_s^0 or two \bar{B}_s^0 hadrons would arise from mixing. Finally, we need to measure the decay time. The B_s oscillations

¹⁴The early observation of this mixing was a surprise because it was only explicable in the Standard Model if the mass of the top quark was very large, and this measurement gave the first indication that the top quark would be so heavy. Here again, we see how precision low-energy measurements can give access to physics at much higher mass scales via the loop diagrams.

¹⁵The leptons in this analysis are restricted to e or μ .

¹⁶ $b \rightarrow c l^- \bar{\nu}_l$, so a negative (positive) lepton arises from the decay of a b (\bar{b}).

¹⁷From conservation of strangeness, the B_s^0 must be produced in association with an \bar{s} quark. If this \bar{s} hadronizes to form a charged kaon, the sign of that kaon identifies the flavour of the B_s^0 .

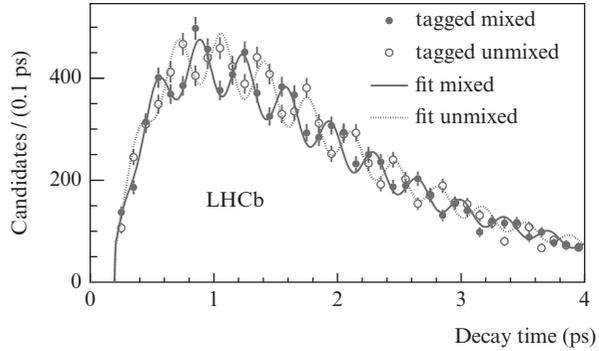


Fig. 10.6 Decay time distributions for B_s^0 for mixed and unmixed events. The suppression of events at low values of the decay time is due to the trigger acceptance. From [104].

¹⁸A mistag occurs when the tagging algorithm arrives at the wrong conclusion. The effect is to dilute the magnitude but not the period of the observed oscillations.

are very rapid, so excellent decay time resolution is essential. The oscillation probability is similar to that derived for kaon oscillations (eqn 10.17) after allowing for the effects of the finite experimental resolution and the probability of a ‘mistag’.¹⁸ We have the characteristic oscillation term $\pm \cos(\Delta M_{B_s} t)$. The sign is positive for mixed flavour at production and decay and negative for the case where the flavour is the same at production and decay. The characteristic oscillations are clearly seen in the data [104] shown in Fig. 10.6.

10.4.5 Regeneration

A further interesting phenomenon related to oscillations, consequent on the very different lifetimes Γ_1 and Γ_2 in the kaon system, is regeneration (the discovery paper is [61]). It is an experimental consideration that must be controlled carefully in order to measure CP violation, as we will see in the next section. Regeneration occurs when we make¹⁹ a beam of pure K_2 and then let that interact by the strong interaction by hitting some material. Because the strong interaction is involved, the K^0 and \bar{K}^0 components of the K_2 must be considered. K^0 and \bar{K}^0 interact differently in nuclear matter. The \bar{K}^0 cross-section is greater than the K^0 cross-section because \bar{K}^0 has a \bar{d} antiquark, which can annihilate with a d quark in a nucleon to produce hyperons (e.g. $\bar{K}^0 + p \rightarrow \Lambda^0 + \pi^+$). There is no equivalent interaction for the K^0 . Regeneration is not the effect of a single collision with the material, it is a quantum-coherent effect—the amplitudes of interactions with many nuclei in the material all sum together. The kaon is not deviated in its path and does not suffer any energy loss.

The consequence of this is that after a K_2 beam passes through an absorber, the amounts of K^0 and \bar{K}^0 will change and the beam will no longer be pure K_2 ; it will contain some amount of K_1 . $K_1 \rightarrow \pi\pi$ decays will be observed again after the absorber. This phenomenon is called *regeneration*. Figure 10.8 shows data from the KTeV experiment at Fermilab,²⁰ which is an example of the regeneration effect.

²⁰To be described in Section 10.5.3.

Regeneration measurements allow the sign of ΔM to be determined. It is found that $\Delta M = M_2 - M_1 = 3.483 \times 10^{-12}$ MeV, i.e. that K_2 is heavier than K_1 .

10.5 *CP* violation (part 1)

In addition to the interesting oscillation effects described up to now, the neutral meson systems also exhibit the fundamental effect of violation of *CP* symmetry. This effect has now been seen in a wide range of places, including decays of charged mesons. There are three main ways in which *CP* violation can be observed, which are listed in Table 10.2 for future reference. There is now strong evidence that the *CP* violation we see is due to a mechanism proposed by Kobayashi and Maskawa connected with the CKM matrix. We will follow the historical route in our description here and discuss the first experimental evidence for *CP* violation, which came from the kaon system, then the Kobayashi–Maskawa mechanism, then briefly other *CP*-violating effects with kaons, which were small and required high-precision experiments. There then followed an extensive period, which is still in progress, where strong *CP*-violating effects in *B*-meson systems became experimentally accessible, leading to tight constraints on the parameters in the CKM matrix.

-
- | | |
|-----|--------------------------------------------------------------------------------|
| (a) | In mixing |
| (b) | In decays |
| (c) | In interference between several decay diagrams leading to the same final state |
-

Table 10.2 Main ways of observing *CP* violation.

10.5.1 Discovery of *CP* violation

The first experimental evidence of *CP* violation came in 1964 when long-lived neutral kaons were observed to decay to two pions [61]. The result of this experiment, which is entirely consistent with what is known today, was the first evidence for *CP* violation and was completely unexpected at the time: the decay $K_2 \rightarrow 2\pi$ was expected to be strictly forbidden by *CP* conservation, as described above. Figure 10.7 shows data from a more recent experiment (NA31 [112]) with high statistics and shows the number of $K^0 \rightarrow \pi^+\pi^-$ decays versus time. The fast exponential component is from the *CP*-conserving decay $K_1 \rightarrow 2\pi$; the constant component that remains after $t \sim 15\tau_1$ shows that $CP = +1$ $\pi^+\pi^-$ states appear to be produced in a region where only $CP = -1$ K_2 states should exist. Shortly after the discovery of *CP* violation in $\pi^+\pi^-$ final states, separate experiments found $\pi^0\pi^0$ states with invariant masses consistent with being kaons produced in the region where only the K_2 (i.e. no K_1) should exist. Both of these effects are violations of *CP* symmetry.

At this point, the question that arises when thinking of the processes as Feynman diagrams is: Where is the *CP* violation happening? It could be either in the Feynman diagrams representing the mixing or in those representing the decay²¹ (or both). The answer, it turns out, is that it is mainly in the mixing in the kaon system.

²¹Types (a) or (b) in Table 10.2.

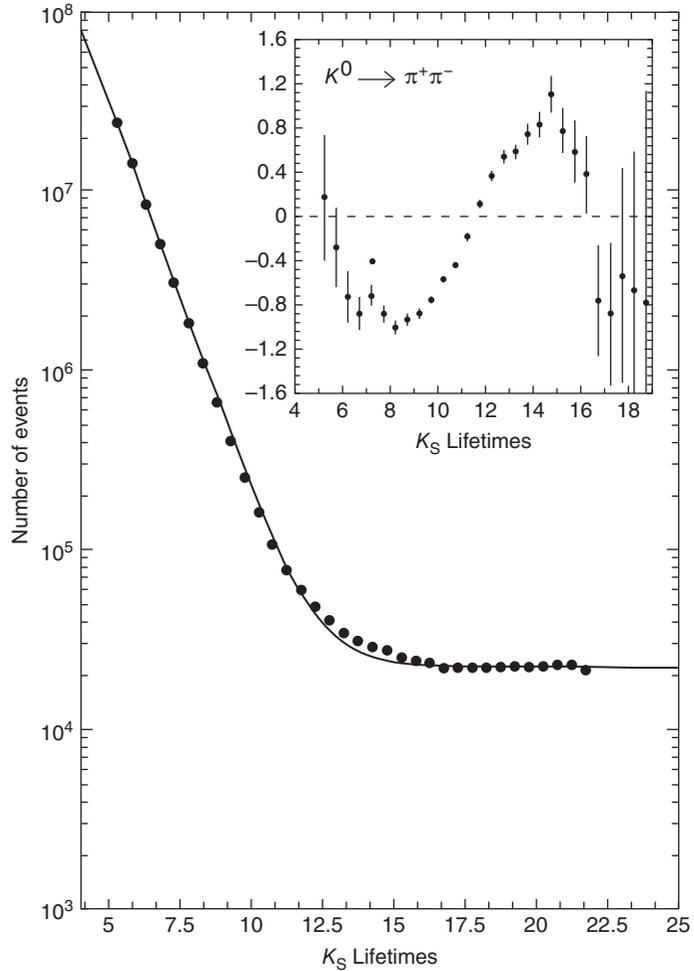


Fig. 10.7 Example of CP violation in $K \rightarrow \pi^+\pi^-$ decays from NA31 [112]. The data are the number of $K \rightarrow \pi^+\pi^-$ decays as a function of time. Below 10 lifetimes, the main feature is CP -conserving K_S decay. Above about 15 lifetimes, the CP violation effect is visible, because the decay rate does not keep falling exponentially. (Interference is possible between these two ways in which the kaons can decay. The inset shows the difference between the data and a fit without the interference. A fit with interference nicely follows the oscillations of the data.)

²²So the formalism we have used up to now is still valid.

²³It turns out that the K_L is the heavier of the two kaon states, so unfortunately the names B_L and K_L do not correspond to each other.

To incorporate the effect of CP violation into the mixing formalism that we have been developing, we introduce two more state labels, K_S and K_L , to represent the physical states that decay with the short and long lifetimes, respectively. If CP were conserved, $|K_S\rangle \equiv |K_1\rangle$ and $|K_L\rangle \equiv |K_2\rangle$, but to include CP violation, we maintain the definitions of K_1 and K_2 as CP eigenstates, defined by eqn 10.1.²² The K_L state represents the component in the beam once all the short-lived kaons have decayed away, which is mostly the $CP = -1$ K_2 , but includes an admixture of a K_1 component, which is what produces the $\pi\pi$ decays. In the other meson systems, the nomenclature is slightly different, although the concept is exactly the same. Since the B mesons all have lots of decay modes, the lifetimes are not very different, and so the states are labelled based on the mass splitting as light²³ B_L and heavy B_H . Returning to Fig. 10.7, the level of CP violation can be characterized by measuring the decay-rate ratio $\eta_{+-} = \Gamma(K_L \rightarrow \pi^+\pi^-)/\Gamma(K_S \rightarrow \pi^+\pi^-)$; the current average of measurements is $\eta_{+-} = (2.232 \pm 0.011) \times 10^{-3}$.

10.5.2 Semileptonic charge asymmetry

The next indication of CP violation was an observation of an asymmetry in the semileptonic decays of the K_L :

$$A_L = \frac{\Gamma(K_L \rightarrow \pi^- l^+ \nu) - \Gamma(K_L \rightarrow \pi^+ l^- \bar{\nu})}{\Gamma(K_L \rightarrow \pi^- l^+ \nu) + \Gamma(K_L \rightarrow \pi^+ l^- \bar{\nu})}. \quad (10.22)$$

The choice of measuring a ratio like this minimizes experimental systematic effects; nevertheless, the experimental set-up needs to have the least amount of material as possible in the path of the particles because the interaction cross sections of π^+ and π^- are different²⁴ and could affect the measurement if too large. The semileptonic asymmetry A_L is measured to be $(3.32 \pm 0.06) \times 10^{-3}$. If CP were conserved, it would be zero. The fact that this asymmetry is present demonstrates that the CP violation is occurring in the mixing of the kaons rather than in their decay. Another way of saying this is that this measurement shows that $\langle K_L | K_S \rangle \neq 0$.

We can formulate the combined effects of mixing and CP violation using the relations in eqns 10.1 between the K_1 , K_2 and the K^0 , \bar{K}^0 , and then expressing K_S and K_L in terms of K^0 and \bar{K}^0 as follows:

$$\begin{aligned} |K_S\rangle &= p|K^0\rangle + q|\bar{K}^0\rangle \\ |K_L\rangle &= p|K^0\rangle - q|\bar{K}^0\rangle \end{aligned} \quad (10.23)$$

where p and q are complex coefficients with $|p|^2 + |q|^2 = 1$. $\langle K_L | K_S \rangle = |p|^2 - |q|^2$, which is zero if CP is conserved. Considering the charge asymmetry A_L again, let the amplitude for $K^0 \rightarrow \pi^- l^+ \nu$ be f , so the amplitude for $\bar{K}^0 \rightarrow \pi^+ l^- \bar{\nu}$ is f^* . It is now possible to show that $A_L = |p|^2 - |q|^2$ and therefore that the non-zero measured value of A_L indicates that the K_S and K_L are not orthogonal to each other, and hence that CP violation is occurring in the mixing.

An alternative way to write the effects of mixing and CP violation is to define K_1 and K_2 in terms of K^0 and \bar{K}^0 with eqns 10.1 as before; and then define K_S and K_L in terms of K_1 and K_2 with a small complex impurity parameter $\tilde{\epsilon}$ as

$$|K_S\rangle = \frac{|K_1\rangle + \tilde{\epsilon}|K_2\rangle}{\sqrt{1 + |\tilde{\epsilon}|^2}} \quad (10.24)$$

$$|K_L\rangle = \frac{|K_2\rangle + \tilde{\epsilon}|K_1\rangle}{\sqrt{1 + |\tilde{\epsilon}|^2}} \quad (10.25)$$

With this definition, $p/q = (1 + \tilde{\epsilon})/(1 - \tilde{\epsilon})$, $A_L = 2 \operatorname{Re}(\tilde{\epsilon})/(1 + |\tilde{\epsilon}|^2) \sim 2 \operatorname{Re}(\tilde{\epsilon})$, and $\eta_{+-} = \tilde{\epsilon}$. From 1964 until around 1999, all the CP -violating effects measured could be characterized by the single parameter $\tilde{\epsilon}$.

Theories were proposed to explain CP violation when it was first discovered. In particular, a new $\Delta S = 2$ ‘superweak’ interaction specific to the kaon system was considered. The CP violation in the kaon system

²⁴For similar reasons to those accounting for the difference in K^0 and \bar{K}^0 interaction cross sections in Section 10.4.5.

was studied extensively over nearly forty years before evidence for CP violation in B^0 systems—which will be discussed shortly—was found. The next sections will outline how CP violation can be accommodated in the Standard Model by introducing a complex phase into the 3×3 CKM quark mixing matrix (see Chapter 7).²⁵

²⁵This requires a third generation and it is interesting to note that this was proposed after CP violation had been observed but long before either of the third-generation quarks had been discovered.

10.5.3 CP violation in K^0 decay

A signature that could be used to gain insight into the mechanism causing CP violation is to look for CP violation in the decay of particles—it was first observed in the decay of K_L particles. If all the CP was occurring in the mixing, then the decays of K_L to $\pi\pi$ should have exactly the same features as the decays of K_S to $\pi\pi$, because, in both cases, it is just K_1 decaying. In particular, the ratio of decays to $\pi^0\pi^0/\pi^+\pi^-$ should be the same for both K_L and K_S . There was an extensive programme of experimental research to measure this, and the results were expressed in terms of the double ratio $R = |\eta_{00}|^2/|\eta_{+-}|^2$, where

$$\eta_{00} = \frac{A(K_L \rightarrow \pi^0\pi^0)}{A(K_S \rightarrow \pi^0\pi^0)}, \quad \eta_{+-} = \frac{A(K_L \rightarrow \pi^+\pi^-)}{A(K_S \rightarrow \pi^+\pi^-)}$$

and $A(K \rightarrow \dots)$ represents the amplitudes of the decays. Unfortunately, most of the $K \rightarrow \pi\pi$ decays produce the pions in a single isospin state ($I = 0$), and so, from isospin symmetry (see Section 5.2.1), the ratio $\pi^0\pi^0/\pi^+\pi^-$ is the fixed value of $\frac{1}{2}$ no matter what the production mechanism of the $\pi\pi$ state is. However, a small fraction of the $\pi\pi$ are made in an isospin $I = 2$ state, which has a different value of the $\pi^0\pi^0/\pi^+\pi^-$ ratio, and so it is possible to use the double ratio R to detect CP violation in decays; the effect is very small, however.

Experimentally, this double ratio was a good quantity to measure since it allowed tricks to cancel systematic errors, for example the use of the same detectors to measure the K_L and K_S . The NA31 experiment at CERN took data alternately in K_L mode (where the target was far upstream to allow the K_S to decay before reaching the experiment) and K_S mode (where a target close to the experiment, with a far less intense proton beam, was used; the K_S target was moved on rails to different positions to reproduce a decay distribution similar to the almost-flat K_L beam to minimize acceptance differences). The KTeV experiment at Fermilab [93] used two simultaneous K_L beams²⁶ with an absorber in one of them to regenerate a K_S beam (its predecessor, the E731 experiment [116], used a similar technique). Figure 10.8 shows the reconstructed decay position distribution along the beam direction z and illustrates the difference in decay distributions from the K_L and K_S . It also illustrates the phenomenon of kaon regeneration. NA48 (a successor to NA31) had both a K_L and a K_S beam, produced from separate proton beams. Both KTeV and NA48 allowed simultaneous measurement of K_L and K_S to remove any small time-varying systematic effects in the detectors. The layout of NA48 is shown in Fig. 10.9.

²⁶In reality one big beam made from a target at $z = 0$ that was collimated into two beams side by side. The regenerator beam was attenuated to avoid a huge rate of decays, which is why no K_L is visible in that beam.

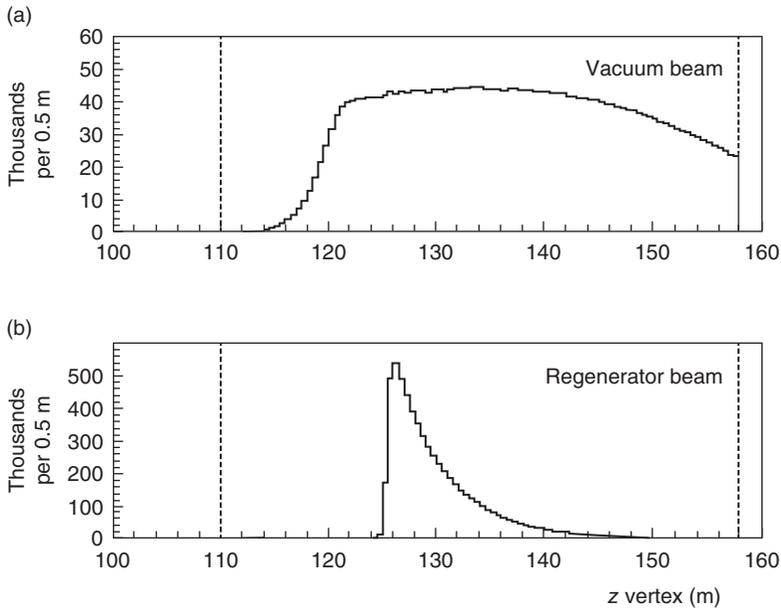


Fig. 10.8 Distribution of decay positions z of $\pi^+\pi^-$ events along the beam direction in the KTeV experiment [93], which used two beams of K_L and a regenerator (a block of material) situated at $z = 125$ m in one of them to produce K_S particles. The regeneration effect is clearly visible and the difference in the decay distributions due to the different K_S and K_L lifetimes is very apparent. The acceptance of the detector varies slowly with z , which produces the non-flat K_L distribution. Decays between $z = 110$ and 158 m are used in the analysis.

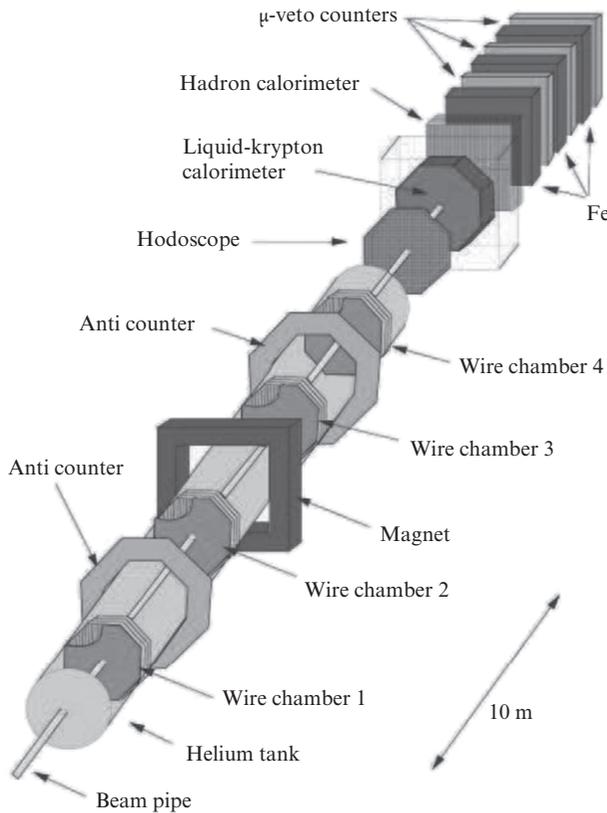


Fig. 10.9 Layout of the NA48 experiment at CERN to measure CP violation in kaons. By comparison of the components in the detector with a collider experiment, the detector components are really very similar, just rolled out in a line rather than in a cylinder, so that the particles encounter the tracking + magnet first, then the calorimetry (electromagnetic, then hadronic), and finally the muon detectors. The kaon experiments are very long and narrow, reflecting the large Lorentz boost to which the ~ 100 GeV kaons are subjected. From [113].

²⁷Liquid argon and lead for NA31, lead glass for E731, caesium iodide crystals for KTeV, and liquid krypton for NA48.

All four of these experiments had extremely high-resolution electromagnetic calorimeters²⁷ capable of measuring high rates of particles, which were vital to separate the $\pi^0\pi^0$ decays from the far more numerous CP -conserving $K_L \rightarrow \pi^0\pi^0\pi^0$ decays. The $\pi^+\pi^-$ decays were measured with a spectrometer (magnet + drift chambers) and backgrounds from three-body decays ($\pi^+\pi^-\pi^0$, $\pi^\pm e^\mp\nu$, and $\pi^\pm\mu^\mp\nu$) were removed by (a) looking at the transverse momentum p_T distribution of the events (most two-particle events that are background have larger p_T , indicating another particle that was missed), (b) checking that the momentum in the spectrometer p was inconsistent with the energy in the electromagnetic calorimeter E (since E/p is close to 1 for electrons), and (c) checking that the reconstructed invariant mass of the two tracks was close to m_K . The experiments also had sophisticated multilevel triggers.

The combined result of all these experiments is $R = 0.9899 \pm 0.0012$. Since this is not consistent with $R = 1$, this is evidence for CP violation in the decay of the K_L particle, or, equivalently, it shows that the K_2 state can decay to two pions directly. This excluded the superweak interpretation of CP violation and was consistent with the CKM model of CP violation, to be described next. To incorporate this into the formalism above, the parameter $\tilde{\epsilon}$ is replaced by two parameters²⁸ ϵ ($\simeq \tilde{\epsilon}$) and ϵ' , where $\epsilon \sim 2 \times 10^{-3}$ characterizes the CP violation in the mixing and $\epsilon' \sim 3 \times 10^{-6}$ characterizes CP violation in the decay (type (b) in Table 10.2).

²⁸It can be shown that $\eta_{+-} = \epsilon + \epsilon'$, $\eta_{00} = \epsilon - 2\epsilon'$, and $R = 1 - 6\text{Re}(\epsilon'/\epsilon)$

10.6 CP violation in the Standard Model

The CKM matrix was introduced in Section 7.3.3 to explain the rotation between quark states $|d\rangle, |s\rangle, |b\rangle$ (flavour eigenstates) produced in strong interactions and the $|d'\rangle, |s'\rangle, |b'\rangle$ states that couple with the W boson. The CKM matrix elements are needed in weak decays involving quarks. As also mentioned in Section 7.3.3, Kobayashi and Maskawa found that by extending the matrix to be a 3×3 matrix, they were able to insert a non-trivial complex phase into it. The presence of the phase, which appears in the transition amplitudes, can cause T violation, since

$$T(e^{-iEt+\delta}) \rightarrow e^{iEt+\delta}$$

and hence, via the CPT theorem, CP violation is expected.

This does not work with a 2×2 matrix, in which the unitarity condition imposes that the complex phases can be removed without affecting any observables, and so the discovery of CP violation along with the work of Kobayashi and Maskawa was an early indication for the third generation of quarks. Products of the CKM elements appear in the meson decay amplitudes, and the phase from the CKM matrix accounts for the CP -violating effects seen so far. The formalism allows insights into the likely magnitude of CP -violating effects in other processes by examining the relation between the elements.

The unitarity of the CKM matrix \mathbf{V} requires that $\mathbf{V}^\dagger \mathbf{V} = 1$. In terms of the individual elements, this²⁹ gives nine relationships:

$$\sum_{i=1,3} |V_{ij}|^2 = 1 \quad (j = 1, 2, 3) \quad (10.26)$$

$$\sum_{i=1,3} V_{ji} V_{ki}^* = \sum_{i=1,3} V_{ij} V_{ik}^* = 0 \quad (j, k = 1, 2, 3; j \neq k) \quad (10.27)$$

The six relations of eqn 10.27, each of which is a sum of three complex numbers, form a ‘unitarity’ triangle in the complex plane. It can be shown that

$$|\text{Im}(V_{km}^* V_{lm} V_{kn} V_{ln}^*)| = |\text{Im}(V_{mk}^* V_{ml} V_{nk} V_{nl}^*)| = J \quad (10.28)$$

irrespective of k, l, m, n , and all six triangles have the same area, $A = \frac{1}{2}J$, independent of any phase convention. J is known as the *Jarlskog invariant*.

Figure 10.10 shows one of the triangles involving the CKM elements responsible for strange and charm (D -meson) decays and one responsible for B decays. The angles of the triangle in Fig. 10.10(b), which represents the unitarity constraint $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$, are large. We will see in Section 10.7.1 that the angles in the unitary triangle determine the magnitude of the CP -violating effects. Therefore, the large angles in the triangle describing B^0 decays correspond to large CP -violating effects. Conversely, the angles in the triangle describing K^0 and D^0 decays are small. The CKM ansatz therefore predicts large CP violation in B -meson systems but very little in K^0 and D^0 systems. In fact, any CP violation in D^0 systems is expected to be almost negligibly small.³⁰ Any CP violation in D^0 decays at a level of more than a few parts in a thousand would be strong evidence for new physics.

A second way to explore whether phenomena beyond the CKM explanation of CP violation exist is to make accurate measurements of the structure of the CKM matrix. By making separate measurements of the angles α, β , and γ shown in Fig. 10.10, and the lengths of the sides of the triangle, we can check whether they are all consistent. Inconsistency could indicate, for example, that the 3×3 matrix is just a part of a larger matrix, or that the CKM formalism is not correct.

To find processes in which CP -violating effects could be present, two things are needed:

- (1) The process must have diagrams involving all three generations of quarks (if this is not true, we could shift the complex phase into the part of the matrix of the missing generation, so it would not appear in the decay amplitudes).
- (2) There needs to be more than one diagram to get to the final state (if there is only one diagram, the phase drops out when calculating $|A|^2$).

$${}^{29}\mathbf{V}^\dagger = (\mathbf{V}^*)^T.$$

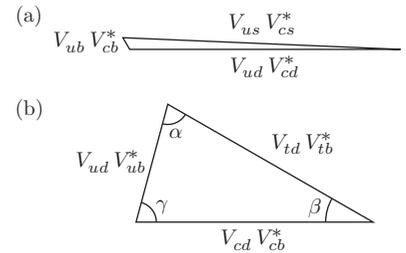


Fig. 10.10 Unitarity triangles responsible for (a) strange and charm decays and (b) B decays. The angles α, β , and γ are also known as ϕ_2, ϕ_1 , and ϕ_3 , respectively.

³⁰The reason for this is that the mixing and decays of K^0 and D^0 involve mostly the first two quark generations; CP violation is observable in the kaon system because of the long lifetime of K_L .

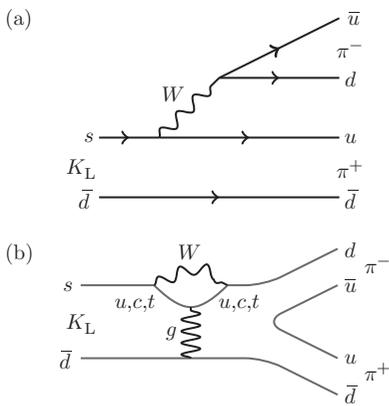


Fig. 10.11 Two Feynman diagrams for K_L decays to two pions. The interference between diagram (a) and diagram (b) (called a ‘penguin diagram’) generates ‘direct’ CP violation (i.e. there is CP violation without mixing).

³¹The parameters p and q can have different values in each of the meson systems, although, from unitarity, $|p|^2 + |q|^2 = 1$ in each case.

The CP violation in the mixing, proceeding via diagrams in Fig. 10.1, satisfies these criteria because there are several such diagrams and some of them involve the t quark inside the loop, and so some of the diagrams involve all three generations. Kaon decay can occur by diagrams of the form shown in Fig. 10.11; the more complicated ‘penguin’ diagrams that can contain quarks from the third generation in the loop provide a mechanism for CP violation in the decay.

10.6.1 Mixing with CP violation

We look again at the mixing in meson decays and derive expressions for the time variation of the states, but now including CP violation as included by eqns 10.23. We have previously used symbols related to the kaons, even though many expressions have been valid for all the meson systems.³¹ For variety (and because we will use the formulae to describe B -meson physics next), we use the symbols for the B mesons here. This involves the substitution from $(K^0, \bar{K}^0, K_1, K_2, K_S, K_L)$ to $(B^0, \bar{B}^0, B_1, B_2, B_L, B_H)$. From eqns 10.23, the light and heavy eigenstates written with B symbols are

$$|B_L\rangle = p|B^0\rangle + q|\bar{B}^0\rangle \quad (10.29)$$

$$|B_H\rangle = p|B^0\rangle - q|\bar{B}^0\rangle$$

Inverting these, we obtain

$$|B^0\rangle = \frac{|B_L\rangle + |B_H\rangle}{2p} \quad (10.30)$$

$$|\bar{B}^0\rangle = \frac{|B_L\rangle - |B_H\rangle}{2q}$$

Equations 10.30 can be used to write the time evolution for a state $|\psi\rangle$ that was created as a B^0 at time $t = 0$:

$$|\psi(t)\rangle = \frac{e^{-\Gamma_B t/2} e^{-iMt}}{2p} \left(e^{i\Delta Mt/2} |B_L\rangle + e^{-i\Delta Mt/2} |B_H\rangle \right) \quad (10.31)$$

We have used the (good) approximation that the B_L and B_H have the same lifetime here. Using eqn 10.29, we can express this as

$$|\psi(t)\rangle = \frac{e^{-\Gamma_B t/2} e^{-iMt}}{2p} \left[e^{i\Delta Mt/2} (p|B^0\rangle + q|\bar{B}^0\rangle) + e^{-i\Delta Mt/2} (p|B^0\rangle - q|\bar{B}^0\rangle) \right] \quad (10.32)$$

$$= e^{-\Gamma_B t/2} e^{-iMt} \left[\cos(\Delta Mt/2) |B^0\rangle + i \frac{q}{p} \sin(\Delta Mt/2) |\bar{B}^0\rangle \right] \quad (10.33)$$

We can now evaluate the probability for a state that was initially a B^0 to be found as a B^0 at time t and similarly the probability that it oscillates into a \bar{B}^0 :

$$\begin{aligned} |\langle B^0 | \psi(t) \rangle|^2 &= e^{-\Gamma_B t} \cos^2(\Delta M t / 2) \\ |\langle \bar{B}^0 | \psi(t) \rangle|^2 &= e^{-\Gamma_B t} \frac{|q|^2}{|p|^2} \sin^2(\Delta M t / 2) \end{aligned} \quad (10.34)$$

The amount of CP violation in the mixing of B mesons is found to be small. In the limit of no CP violation, $|p|/|q| = 1$ and³² B_L and B_H would be orthogonal.

³²The relationship

$$\langle B_L | B_H \rangle = |p|^2 - |q|^2$$

carries over from the kaons

10.7 CP violation (part 2)

The CKM formalism for CP violation with one complex phase δ can be easily made to fit the experimental CP violation effects described up to now, and indeed the value of δ is not strongly constrained by these. One reason is that strong-force (QCD) effects on the theoretical prediction of ϵ' are difficult to calculate,³³ so it is not easy to relate it to δ . However, the CKM formalism gives predictions (as a function of the assumed value of δ) of CP violation in other systems, and, owing to the values of the CKM elements, some of the effects in the B -meson systems can be large. Although the kaons have the advantage that the K_S and K_L have very different lifetimes, there are several tricks for measuring CP violation with B mesons that do not have this feature.

³³It is a complex theoretical computation and there are several large terms that nearly cancel to give a small overall number, which consequently has a large uncertainty.

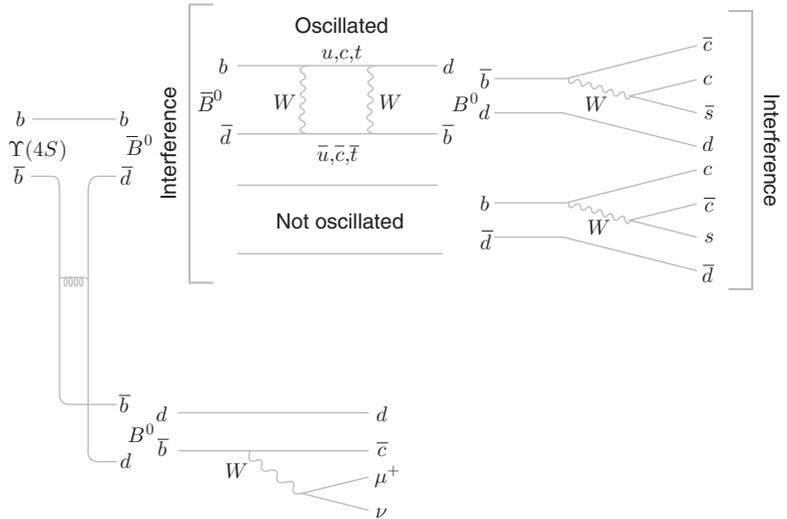
As stated earlier, we need to find a situation where (a) all three quark generations can be involved and (b) several routes lead to the final state, so that when we add the amplitudes together and square to get the observable quantity, there is a dependence on the complex phase in the CKM matrix.

10.7.1 CP violation in time-dependent asymmetries

The first technique that exploits this is elegant and gives a large effect, called the time-dependent asymmetry. It is our first example of type (c) CP violation from Table 10.2. It has been studied in detail at facilities called B-factories and involves making a $B^0 \bar{B}^0$ pair from the decay of an $\Upsilon(4S) b\bar{b}$ meson, which we will look at shortly. It can also be studied at the LHCb, experiment which we will also examine later in this chapter. A final state that is a CP eigenstate is required, one to which both a B^0 and a \bar{B}^0 can decay; a good example of this is $B^0 \rightarrow J/\psi K_S$. A schematic of the entire process³⁴ is shown in Fig. 10.12. It turns out that the CP violation in the mixing of B^0 (in the box diagrams in Fig. 10.4) is very small, but the final state has two routes by which it can be made from a B^0 : one in which the B^0 oscillates to a \bar{B}^0 before decaying and one

³⁴Although the K_S is needed to produce a CP final state, we can draw this in the Feynman diagram as either an $s\bar{d}$ or a $d\bar{s}$; although there is CP violation in the mixing of the kaons, it is small compared with the main source of CP violation in this technique and can be neglected.

Fig. 10.12 Schematic of the processes involved in the time-dependent asymmetry measurements. The $\Upsilon(4S)$ produces a pair of B mesons (shown left). One decays in a way that tags whether there is a b or \bar{b} quark in the meson (shown below), the other decays to a CP eigenstate that is accessible from either the B^0 or \bar{B}^0 (shown above). Provided there are two routes with similarly large amplitudes to get to the same final state, interference between them can reveal the complex phase in the CKM matrix.



where it does not. The amplitudes for these two parts of the decay will interfere, and the sensitivity to the angle δ in the CKM matrix is large.

Considering decays of B mesons to states of definite CP (labelled as f_{CP}), we define the amplitude for $B^0 \rightarrow f_{CP}$ as A_f and the amplitude for a \bar{B}^0 decay to the same state as \bar{A}_f . It is conventional to define the parameter

$$\lambda_f = \frac{\bar{A}_f q}{A_f p} \tag{10.35}$$

From eqn 10.33, we can write down the amplitude representing the decay of a B^0 state, with wavefunction $\psi(t)$, to the f_{CP} state:³⁵

$$\langle f | \psi(t) \rangle = e^{-\Gamma t/2} e^{-iMt} \left[A_f \cos(\Delta Mt/2) + i \frac{q}{p} \bar{A}_f \sin(\Delta Mt/2) \right]$$

Using the definition of λ_f from eqn 10.35, this can be rewritten as

$$\langle f | \psi(t) \rangle = A_f e^{-\Gamma t/2} e^{-iMt} [\cos(\Delta Mt/2) + i \lambda_f \sin(\Delta Mt/2)] \tag{10.36}$$

Similarly, the amplitude representing the decay of a \bar{B}^0 state, with wavefunction $\bar{\psi}(t)$, to the same f_{CP} state is

$$\langle f | \bar{\psi}(t) \rangle = \bar{A}_f e^{-\Gamma t/2} e^{-iMt} \left[\cos(\Delta Mt/2) + \frac{i}{\lambda_f} \sin(\Delta Mt/2) \right] \tag{10.37}$$

Assuming that $|A_f|^2 = |\bar{A}_f|^2$ as expected in the Standard Model, and that $|p| \simeq |q|$ because the CP violation in the mixing is small, we can evaluate the decay rates as (Exercise 10.3)

$$\begin{aligned} |\langle f | \psi(t) \rangle|^2 &= |A_f|^2 e^{-\Gamma t} [1 - 2 \text{Im}(\lambda_f) \cos(\Delta Mt/2) \sin(\Delta Mt/2)] \\ |\langle f | \bar{\psi}(t) \rangle|^2 &= |A_f|^2 e^{-\Gamma t} [1 + 2 \text{Im}(\lambda_f) \cos(\Delta Mt/2) \sin(\Delta Mt/2)] \end{aligned} \tag{10.38}$$

³⁵In the following equations, f_{CP} is shortened to f .

We define the CP -violating asymmetry in the usual way,

$$A_{CP} = \frac{|\langle f|\bar{\psi}(t)\rangle|^2 - |\langle f|\psi(t)\rangle|^2}{|\langle f|\bar{\psi}(t)\rangle|^2 + |\langle f|\psi(t)\rangle|^2} \quad (10.39)$$

and, substituting from eqns 10.38, we find

$$A_{CP} = \text{Im}(\lambda_f) \sin(\Delta Mt) \quad (10.40)$$

The $J/\psi K_s^0$ state is a CP eigenstate (with eigenvalue $\eta_{CP} = -1$), so is suitable for a time-dependent asymmetry analysis. We can use the CKM matrix to predict the phase factors with or without mixing via the box diagram (see Fig. 10.12). In the box diagrams, the important contribution comes from exchange of virtual top quarks owing to the large value of V_{tb} . Note that the asymmetry depends on the phase of λ_f . The phase factors in the box diagram come from the CKM elements and contribute $V_{tb}^2 V_{td}^2$. Since V_{tb} is real, the non-zero phase arises from the factor of V_{td}^2 . Hence the amplitude of the CP -violating term is given by $A_{CP} = \text{Im}(V_{td}^2)$. Now if we look at the unitarity triangle (see Fig. 10.10), we can see that the phase of $V_{td} = \beta$. Hence $A_{CP} = \sin 2\beta$.³⁶

³⁶A similar analysis for the case of $J/\psi K_L^0$ gives a similar result with the sign changed.

10.7.2 B-factories

Since the B -meson system is expected to be a prolific source of physics, many recent studies of B mesons, including the time-dependent asymmetries, have been made at two ‘B-factories’. These are high-luminosity e^+e^- colliders, KEK-B in Japan (with the Belle detector) and PEP-II (and the BaBar experiment) in the USA, operating at a centre-of-mass energy of 10.58 GeV, the mass of the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ is a $C = -1$ state that decays almost entirely to $B_d \bar{B}_d$ pairs with a branching ratio of 49% to $B^0 \bar{B}^0$.³⁷ Studies at e^+e^- B-factories continue with the new Belle-2 experiment at the upgraded Super KEK-B accelerator.

Both machines are ‘asymmetric’ in that the laboratory energies of the electron and positron beams are different—specifically, at PEP-II, the low-energy ring operates at 3.1 GeV and the high-energy ring at 9.0 GeV. The asymmetry between the energies of the two rings means that the centre-of-mass system (CMS) moves at a velocity of $\beta = 0.5$ in the laboratory and the resulting Lorentz boost allows resolution of the B decay vertices, which would otherwise not be resolved if mesons were produced at rest in the CMS. This is illustrated in Fig. 10.13. Nevertheless, the detectors must have extremely good vertex resolution, since the average distance from production to decay, $\beta\gamma c\tau_B$, is only about 260 μm even with the Lorentz boost.

Because $B^0 \bar{B}^0$ pairs from the decay of $\Upsilon(4S)$ are produced essentially at rest in the CMS, and the $\Upsilon(4S)$ is a state of definite (odd) C , the initial state is

$$\Upsilon(4S)_{C=-1} \rightarrow \sqrt{\frac{1}{2}} [|B^0(\mathbf{p})\bar{B}^0(-\mathbf{p})\rangle - |B^0(-\mathbf{p})\bar{B}^0(\mathbf{p})\rangle]$$

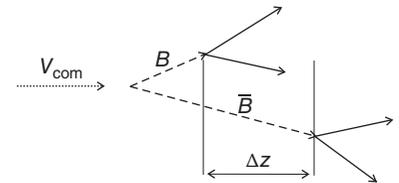


Fig. 10.13 The Lorentz boost due to the finite velocity of the CMS in an asymmetric B-factory allows the resolution of the B -meson decay vertices. The distance between the two decay vertices allow the time between the two B^0 decays to be determined.

³⁷ B_s^0 and \bar{B}_s^0 are too massive to be produced from Υ decays.

where \mathbf{p} is the CMS momentum of one of the B s. The oscillations that occur after production are quantum-correlated: since the state must be odd under the interchange of the two mesons, they cannot both be B^0 or \bar{B}^0 simultaneously—they must oscillate together. The coherence is lost once one of the mesons has decayed and oscillations can be observed—the decay of one meson starts the clock for the time-dependent oscillations of the other. Experimentally, the time $\Delta t = t_1 - t_2$ is determined from the distance between the decay vertices, as sketched in Fig. 10.13.

The scheme of the processes that are selected in the detector to measure the time-dependent asymmetry is shown in Fig. 10.12, which we have partly discussed already. We consider events that decay as $\Upsilon(4S) \rightarrow B^0 \bar{B}^0 \rightarrow f_{CP} f_{\text{tag}}$, where f_{CP} is a CP eigenstate (shown in the top part of Fig. 10.12) and f_{tag} is a B meson that is tagged as either B^0 or \bar{B}^0 through a semileptonic decay as shown in the bottom part of Fig. 10.12. Let $n(B^0)(n(\bar{B}^0))$ be the number of tagged $B^0(\bar{B}^0)$ events observed as a function of Δt . We can redefine the asymmetry A_{CP} from eqn 10.39 in terms of event numbers as

$$A_{CP} = \frac{n(B^0) - n(\bar{B}^0)}{n(B^0) + n(\bar{B}^0)} \quad (10.41)$$

Figure 10.14 shows the numbers of tagged events (a, b) and A_{CP} (c, d) from the Belle experiment [48] as functions of Δt for the processes $B \rightarrow J/\psi K_S$ (a, c) and $B \rightarrow J/\psi K_L$ (b, d). The dataset from Belle has 700 million $B^0 \bar{B}^0$ events. Similar results are also obtained at the other B-factory by the Babar experiment at SLAC with a sample of 460 million

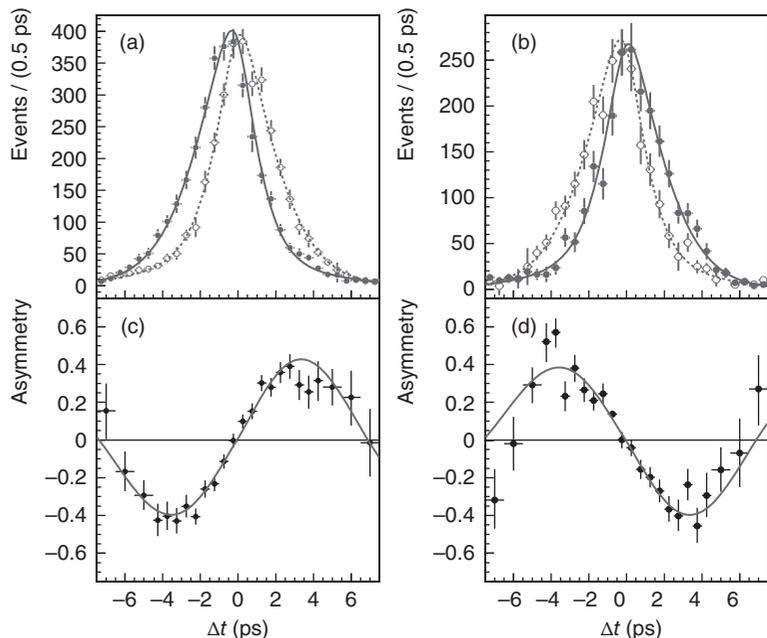


Fig. 10.14 Data from the Belle experiment: (a, b) numbers of events with a B^0 and with a \bar{B}^0 tag as a function of Δt and (c, d) the time-dependent asymmetry. The data are shown separately for the $CP = -1$ (e.g. $J/\psi K_S$) (a, c) and $CP = +1$ ($J/\psi K_L$) (b, d). From [48].

$B^0\bar{B}^0$ events. It can be deduced from these and other measurements that $\sin 2\beta = 0.671 \pm 0.023$.

The B-factories also studied time-dependent asymmetries with other final states. The $B \rightarrow J/\psi K_S$ described here is particularly clean, with diagrams that are easy to calculate theoretically being dominant. This mode is an example of a diagram where the quark-level decay is $b \rightarrow \bar{c}c\bar{s}$, and gives access to the angle β . Measurements of the other angles of the unitarity triangle are more involved but can be done. For example, the angle α can be determined from the time-dependent analysis of $B^0 \rightarrow \pi^+\pi^-$ decays, but this is more challenging because of the small branching ratio for this decay mode and backgrounds from other decay modes. In addition, the theoretical analysis of this mode is more complicated.

There are about eight combinations overall of quark decays that can in principle be studied with different final states. Highly significant (over 5σ) CP -violating effects have been observed in the following final states: $J/\psi K$, $\eta' K$, ϕK , $f^0 K$, $K^+K^-K_S$, $\pi^+\pi^-$, $\psi\pi^0$, D^+D^- , and $D^{*+}D^{*-}$. In some of these cases, there are more than two particles in the measured final state, which in general do not have definite CP , and a certain region of the Dalitz plot (see Chapter 2, page 36, for an example) must be selected to isolate the decays in the CP eigenstate.

10.7.3 LHCb detector

The B-factories using the $\Upsilon(4S)$ can produce B_d mesons but they are below threshold for producing B_s mesons. Hadron colliders offer a prolific source of B mesons, including the B_s mesons, and this will be discussed in this section. The angle γ can be determined most cleanly from measurements of direct CP violation in $B^\pm \rightarrow D^0K^\pm$ decays, or from effects involving the B_s mesons.

The LHCb experiment [99] is a dedicated B -physics facility at the LHC, established following the success in studying B physics at the Tevatron experiments. Because of their relatively small mass compared with the CMS energy, b hadrons are produced predominantly along the beam directions, in what is called the ‘forward direction’. One side of an LHC interaction region has been instrumented with detectors specifically optimized for (a) good spatial resolution (which gives good decay time resolution) to distinguish and measure B decays, (b) excellent particle identification (e.g. in distinguishing $B \rightarrow \pi^+\pi^-$ decays from background $B \rightarrow K^+\pi^-$), and (c) a very efficient trigger for selecting B events.

The spatial resolution is achieved with a silicon vertex detector (see Section 4.6.2).³⁸ Because of the geometry of measuring particles in the forward direction, the LHCb silicon vertex detector fits inside the vacuum beam pipe. The detectors are split in two parts so that they can be moved closer to (further away from) the beam during normal operation (filling the accelerator with particles) with motors to avoid exposure to too much radiation. Thin foils surround the vertex detectors to provide shielding, since otherwise the pickup from the moving charges in the beam nearby would swamp the signals of the particles being detected.

³⁸The large boost at LHC helps improve the proper-time resolution.

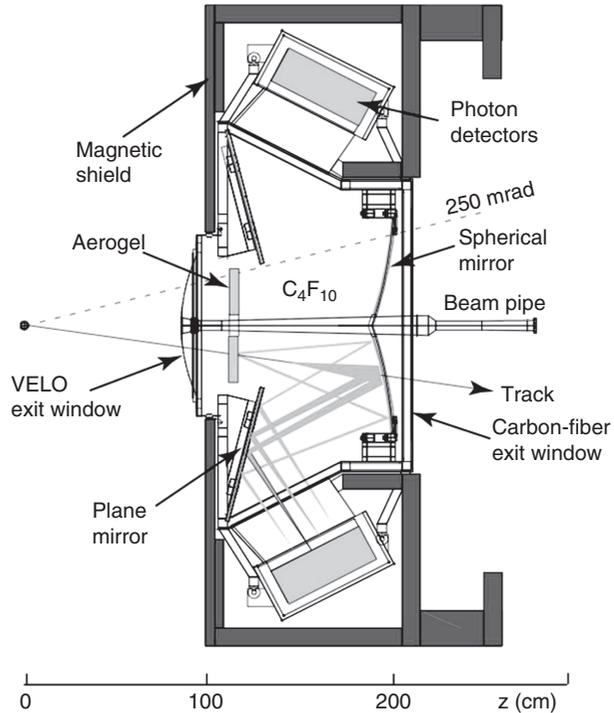


Fig. 10.15 The LHCb RICH detector. Čerenkov photons created by charged particles with speed $\beta > 1/n$ are focused onto the surface of the photon detectors by a series of mirrors. The radiator C_4F_{10} is selected because of its high refractive index among the inert gases. From [99].

The particle identification is mainly needed for K/π separation. This is achieved over a wide range of momenta using two ring imaging Čerenkov (RICH) detectors, one of which is shown in Fig. 10.15. The Čerenkov effect is described in Section 4.3.4; when a particle travelling at speed β is sufficiently fast that $\beta > 1/n$, light is emitted at an angle θ_C given by $\cos \theta_C = 1/(n\beta)$, where n is the refractive index of the medium. In a RICH, the locations of the Čerenkov photons are measured and used to fit a cone around the particle trajectory to determine θ_C . In combination with the momentum measurement from the magnetic spectrometer, this can in principle determine the particle mass.³⁹

³⁹In practice, we do not want to measure the mass but simply to provide separation between particles of different masses.

The LHCb trigger consists of hardware and software levels following the multilevel scheme described in Section 4.10. The hardware trigger uses simple algorithms based on energy deposition in the calorimeters and hits in the muon chambers to reduce the rate to 1 MHz. All the data are read out at this frequency and a large computer farm performs more sophisticated calculations to reduce the rate to 4 kHz, and the data that pass this higher-level trigger are retained on permanent storage for subsequent analysis.

10.8 LHCb measurements

LHCb has made time-dependent asymmetry measurements that are as precise as the B-factory measurements (Section 10.7.2). The decay

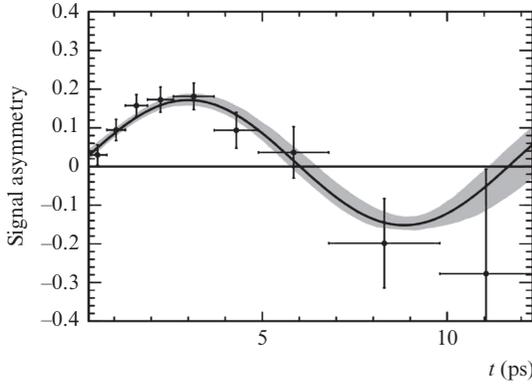


Fig. 10.16 Time-dependent CP -violating asymmetries observed by the LHCb experiment in the decay mode $B^0 \rightarrow J/\Psi K_s^0$. From [102].

$B^0 \rightarrow J/\Psi K_S$ can be reconstructed from the decays $J/\Psi \rightarrow \mu^+ \mu^-$ and $K_S \rightarrow \pi^+ \pi^-$. These particular final states are the simplest to detect in hadronic collisions.⁴⁰ The decay time of the B^0 is reconstructed from the separation between the primary and secondary vertex (measured by the silicon detectors) and the momentum of the B^0 . The flavour of the ‘non-signal B^0 ’ is determined by specific decays, including semileptonic decays to electrons or muons. The time-dependent asymmetry (eqn 10.41) is shown [102] in Fig. 10.16. This is the type (c) CP violation from Table 10.2.

In contrast to the kaon system, CP violation due to mixing alone in the B system does not occur at a high level. This is the type (a) CP violation from Table 10.2 and is characterized by $|q/p| \neq 1$. This can be measured by a time-independent semileptonic asymmetry, which implies that $|q/p| = 1.0002 \pm 0.0028$.

Direct CP violation (type (b) from Table 10.2) has been observed as a difference in the decay rates of neutral B mesons to CP -conjugate final states. For example, for the CP -conjugate decays $B^0 \rightarrow K^+ \pi^-$ and $\bar{B}^0 \rightarrow K^- \pi^+$, the asymmetry parameter is measured to be

$$A(B^0 \rightarrow K^+ \pi^-) = \frac{\Gamma_{K^- \pi^+} - \Gamma_{K^+ \pi^-}}{\Gamma_{K^- \pi^+} + \Gamma_{K^+ \pi^-}} = -0.080 \pm 0.008 \quad (10.42)$$

A similar asymmetry has been measured in the decay of charged B mesons to the $K \rho^0$ final state: $A_{K \rho^0} = 0.37 \pm 0.1$.⁴¹ This shows that CP effect can be measured in charged mesons and is not limited to the four main neutral meson systems.

The first evidence for CP violation in B_s^0 decays [101] uses the mode $B_s^0 \rightarrow K^- \pi^+$ and is an example of direct CP violation. This decay mode has a very small branching ratio and therefore good background rejection is essential. This is achieved using the excellent K/π separation and mass resolution. The measurement is $A(B_s \rightarrow K^- \pi^+) = 0.27 \pm 0.07$, defined similarly to eqn 10.42.

⁴⁰ A disadvantage is that the branching ratio of B^0 to this final state is only $\sim 10^{-3}$.

⁴¹ Similar asymmetry parameters for over two hundred final states of charged and neutral B mesons have been measured and only a few of these, including the two mentioned, are statistically different from zero.

Chapter summary

- Quantum mechanics allows for the possibility of quark–antiquark oscillations.
- These oscillations have been clearly observed in the K^0 , D^0 , B^0 , and B_s^0 mesons. The oscillations can be accounted for by loop Feynman diagrams called ‘box diagrams’.
- CP violation has been observed in K^0 , B^0 , and B_s^0 mesons.
- CP violation can be classified as arising from direct decays, mixing, and interference between decays with and without mixing.
- CP violation in the quark sector can be accommodated in the Standard Model with the introduction of a complex phase in the CKM matrix.
- Precision CP violation measurements are a very sensitive probe of new physics at very high mass scales.

Further reading

- Sozzi, M. S. (2008). *Discrete Symmetries and CP violation: From Experiment to Theory*. Oxford University Press.
- Gershon, T. and Nir, Y. (2012). CP violation in the quark sector. *Phys. Rev. D*, **86**, 010001.
- LHCb Collaboration and A. Bharucha, A. et al. (2013). Implications of LHCb measurements and future prospects. *EPJ C*, **73**, 2373.

Exercises

(10.1) Starting from the definitions of p and q in eqns 10.23, show that $A_L = |p|^2 - |q|^2$.

(10.2) In the NA48 experiment shown in Fig. 10.9, the four photons from $\pi^0\pi^0$ decays are reconstructed in the liquid-krypton calorimeter, yielding 12 quantities E_i , x_i , y_i for $i = 1, \dots, 4$, where E_i is the energy of the i th photon and x_i and y_i are the positions in each of the two directions transverse to the beam. If the mass of the kaon, m_K , is inserted as input, show that the distance in front of the calorimeter z along the beam direction may be reconstructed as

$$z^2 = \sum_{i,j=1}^4 E_i E_j \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{m_K^2}$$

Hint: First consider a $\pi^0 \rightarrow \gamma\gamma$ decay with the y axis perpendicular to the plane of the decay, and discard non-leading terms in x^2/z^2 .

(10.3) Starting from eqns 10.36 and 10.37, derive the CP -violating asymmetry (eqn 10.40).

(10.4) Figure 10.12 shows the processes that are important in the study of time-dependent asymmetries for $B \rightarrow J/\psi K_S$. This technique can also be used (with more subtleties) for other modes. Draw a

similar set of diagrams to show how the process $B \rightarrow \pi^+\pi^-$ proceeds. Penguin diagrams (similar to Fig. 10.11) are non-negligible in this mode; add an example penguin to your diagram. Another mode that exhibits time-dependent asymmetries is $B \rightarrow \phi K_S$; draw the relevant set of diagrams for this mode.

- (10.5) For the decay $B_S \rightarrow \pi^+K^-$, draw a tree diagram and a penguin diagram for this decay. Show that the tree diagram is proportional to $V_{ub}^*V_{ud}$,

and deduce the CKM matrix elements upon which the penguin diagram depends. You should find that the flavour inside the dominant penguin is charm. Repeat for the $\bar{b}d$ decay $B^0 \rightarrow K^+\pi^-$ and notice that the arrangement of the lines in these two decays is very similar. It turns out that because one can show that $\text{Im}(V_{ub}^*V_{ud}V_{cb}V_{cd}^*) = -\text{Im}(V_{ub}^*V_{us}V_{cb}V_{cs}^*)$, the amount of direct CP violation in each of these two decays is the same, which is a result that has been experimentally tested [101].

11

Neutrino oscillations

11.1 Introduction	306
11.2 Neutrino states	307
11.3 Two-flavour oscillations	308
11.4 Evidence for neutrino oscillations	310
11.5 Three (or more)-flavour oscillations	323
Chapter summary	331
Further reading	331
Exercises	331

11.1 Introduction

In the minimal Standard Model (SM), the neutrinos are assumed to be massless. However, as the neutrino masses are not protected by any gauge symmetry (unlike the photon), it is easy to extend the SM to accommodate neutrino masses. From an experimental perspective, it is much easier to detect small mass differences between different neutrino mass eigenstates than to measure their absolute masses. This is because small mass differences combined with mixing will cause one flavour of neutrino to oscillate into other flavours, which is an effect that can be measured. This chapter starts with a very brief review of the determination of upper limits on the neutrino masses and then gives an explanation of the theory of neutrino flavour oscillations. We begin with the simple case of two-neutrino oscillations, since this brings out the essential features with minimal complications. We then review the experimental evidence for neutrino oscillations. Three-generation mixing is very interesting, since it allows for the possibility of CP violation in the neutrino sector, and the mathematical treatment will be given before showing the recent experimental evidence for a non-zero value of the mixing angle between first- and third-generation neutrinos. It is possible that CP violation in the neutrino sector can explain the observed matter–antimatter asymmetry in the Universe, and this will be discussed in Section 11.5.4.

11.1.1 Neutrino masses

The oscillation experiments measure mass differences between the different types of neutrinos, but they are not sensitive to the absolute mass of any neutrino (see Section 11.3). In principle, neutrino masses can be measured from the endpoint(s) of decay spectra. For example, the β decay of tritium, ${}^3\text{H} \rightarrow {}^3\text{He} + e^- + \bar{\nu}_e$, has an endpoint at the Q value of this reaction of 18.6 keV, assuming $m_{\nu_e} = 0$.¹ The effect of a non-zero value of m_{ν_e} would be to create a distortion in the spectrum near the endpoint. Similar studies have been performed using μ and τ decays. So far, only upper limits have been obtained [115]. Limits on the sum of the neutrino masses can also be obtained from cosmological arguments, since finite-mass neutrinos could contribute to the matter density of the

¹Tritium is the best choice here because of the very low Q value.

Universe and affect structure formation in the early Universe. The upper limits depend on which theoretical assumptions are made, but the current upper limit for the sum of the neutrino masses (for all generations) is around 1 eV [115]. A more direct but less precise limit on the mass of the electron neutrino can be derived from the observations of neutrinos from the supernova SN1987A (see Exercise 11.4).

Neutrinos could be of Dirac or Majorana type (see Chapter 6). If they are of Majorana type, this would imply that a neutrino could be its own antiparticle, which would allow neutrinoless double β decay, ${}^X Y \rightarrow {}^X Z + 2e^-$. The Standard Model background to its detection would be normal double β decay, ${}^X Y \rightarrow {}^X Z + 2e^- + 2\bar{\nu}_e$. If we consider the combined energy of the two electrons, the Standard Model background would produce a continuous spectrum as opposed to the peak at the endpoint expected for the neutrinoless double β decay. Several experiments are now looking for this signal, for example the SNO+ experiment is using the isotope ${}^{130}\text{Te}$, which is a double β emitter. The experiment uses liquid scintillator in the SNO detector (previously filled with heavy water for solar neutrino studies). The scintillation process gives a larger number of photons than from the Čerenkov process (which generates the photons in water). Hence the energy resolution for low-energy electrons is significantly better using liquid scintillator compared with heavy water.

11.2 Neutrino states

We know from studying the weak interaction that particle mass states need not be the same as the weak-interaction states. The two types of state are connected by the Cabbibo rotation matrix. A similar phenomenon occurs for neutrino states. If neutrinos have a small mass, the flavour (i.e. weak-interaction) eigenstates are related to the mass eigenstates via a CKM-like matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (11.1)$$

The flavour states ν_e , ν_μ , and ν_τ propagate in space-time as linear combinations of the mass eigenstates ν_1 , ν_2 , and ν_3 . The mixing matrix \mathbf{U} is known as the PMNS (Pontecorvo–Maki–Nakagawa–Sakata) matrix and, as will be discussed in Section 11.5.2, can be generalized to an arbitrary number N of flavours. For the general case of $N \geq 3$, the elements $U_{\alpha i}$ may be complex. The mixing causes transitions between different flavours such as ν_μ and ν_τ . As will be shown below, the flavour transition probabilities $P(\nu_\alpha \rightarrow \nu_\beta)$ depend on the differences in masses between the mass eigenstates and are oscillatory functions of L/E , where L is the distance a neutrino of energy E has travelled from the source to a detector.

11.3 Two-flavour oscillations

Most of the original evidence for neutrino oscillations came from the study of muon neutrinos originating at the top of the atmosphere and electron neutrinos from the Sun. Since the initial neutrino flavours are different and the respective ranges of L/E are very different, the results were usually analysed assuming only two neutrino flavours. It turns out, as will be discussed in Section 11.5.2, that this gives a good description of the main features of neutrino oscillation phenomena. Since the expressions for the transition ('oscillation') probabilities for three (or more) neutrino flavours are somewhat complicated, the physics of two-flavour oscillations will be discussed first. This has the additional advantage that the underlying physics is more transparent. A discussion of three-flavour oscillations is given in Section 11.5.2.

The derivation of two-flavour neutrino oscillation probabilities is similar to the derivation of the $K^0-\bar{K}^0$ oscillation probabilities of Section 10.3, with the important difference that it proceeds in the laboratory frame, rather than the centre-of-mass.

With only two flavours, neutrino mixing can be described by one ('mixing') angle. The two flavour eigenstates ν_α and ν_β (e.g. $\alpha, \beta = e, \mu$) are linear combinations of mass eigenstates ν_1 and ν_2 :

$$\begin{aligned} |\nu_\alpha\rangle &= \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle \\ |\nu_\beta\rangle &= -\sin\theta |\nu_1\rangle + \cos\theta |\nu_2\rangle \end{aligned} \quad (11.2)$$

Consider a neutrino of flavour α with momentum p created in a weak interaction at $t = 0$. The initial state is $|\psi(0)\rangle = |\nu_\alpha(0)\rangle$, which can be expressed in terms of the mass eigenstates² as

$$|\psi(0)\rangle = \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle$$

At some time t later, the ν_1 and ν_2 wavefunctions have evolved:

$$\begin{aligned} |\psi(t)\rangle &= \cos\theta |\nu_1\rangle e^{-iE_1 t} + \sin\theta |\nu_2\rangle e^{-iE_2 t} \\ &= \left(\cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle e^{-i(E_2 - E_1)t} \right) e^{-iE_1 t} \end{aligned} \quad (11.3)$$

where E_1 and E_2 are the energies of the ν_1 and ν_2 . Now

$$\begin{aligned} E_i &= \sqrt{p^2 + m_i^2} \\ &\simeq p \left(1 + \frac{m_i^2}{2p^2} \right) \end{aligned}$$

where, because m_1 and m_2 are undoubtedly very small, the approximation is good for all practical values of p and $p = E$. Ignoring the common

²This treatment follows the simplified methods generally used in textbooks. There are many subtleties in a more correct analysis, but the final results are unchanged (see Akhmedov and Smirnov in Further Reading). We assume that we have a superposition of neutrinos with the same momentum but different energies. In Exercise 11.1, we investigate the effect of changing this assumption. A more general treatment involves the consideration of wavepackets with a finite spread in momentum. A possible objection to our treatment is that it appears to violate conservation of energy since we have neutrino states with (very slightly) different energies. This issue is investigated in Exercise 11.2.

phase factor $e^{-iE_1 t}$ (which will later be cancelled by its conjugate) and rearranging eqn 11.3,

$$\begin{aligned} |\psi(t)\rangle &= \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle e^{-i(m_2^2 - m_1^2)t/2E} \\ &= \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle e^{-i\Delta m^2 t/2E} \\ &= \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle e^{-i\phi} \end{aligned} \quad (11.4)$$

The phase difference $\phi = \Delta m^2 t/2E$ between the ν_1 and ν_2 components of $|\psi(t)\rangle$ depends on $\Delta m^2 = m_2^2 - m_1^2$, the difference between the squared masses of ν_1 and ν_2 , and t/E .

The amplitude of $|\psi\rangle$ at t is given by eqn 11.4 in terms of $|\nu_1\rangle$ and $|\nu_2\rangle$. It can be re-expressed in terms of $|\nu_\alpha\rangle$ and $|\nu_\beta\rangle$ by inverting eqn 11.2:

$$\begin{aligned} |\nu_1\rangle &= \cos\theta |\nu_\alpha\rangle - \sin\theta |\nu_\beta\rangle \\ |\nu_2\rangle &= \sin\theta |\nu_\alpha\rangle + \cos\theta |\nu_\beta\rangle \end{aligned}$$

and hence

$$|\psi(t)\rangle = (\cos^2\theta + \sin^2\theta e^{i\phi})|\nu_\alpha\rangle + \cos\theta \sin\theta (e^{i\phi} - 1)|\nu_\beta\rangle$$

The probability of observing a ν_β at a time t is therefore

$$P(\alpha \rightarrow \beta) = |\langle \nu_\beta | \psi(t) \rangle|^2 \quad (11.5)$$

$$= \cos^2\theta \sin^2\theta (e^{i\phi} - 1)(e^{-i\phi} - 1) \quad (11.6)$$

$$= \cos^2\theta \sin^2\theta (2 - 2\cos\phi) \quad (11.7)$$

$$= 4\cos^2\theta \sin^2\theta \sin^2\left(\frac{1}{2}\phi\right) \quad (11.8)$$

$$= \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 t}{4E}\right) \quad (11.9)$$

Likewise, the probability of observing a ν_α as a ν_α at t is

$$\begin{aligned} P(\alpha \rightarrow \alpha) &= |\langle \nu_\alpha | \psi(t) \rangle|^2 \\ &= 1 - \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 t}{4E}\right) \end{aligned}$$

Since neutrinos travel at speed³ c , the probability that a ν_α of energy E is observed as a ν_β at a distance L from a source is therefore

$$P(\alpha \rightarrow \beta) = \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \quad (11.10)$$

and the probability that the ν_α is observed as a ν_α at L is

$$P(\alpha \rightarrow \alpha) = 1 - \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \quad (11.11)$$

Note that $P(\alpha \rightarrow \alpha) + P(\alpha \rightarrow \beta) = 1$, as required by unitarity.⁴

³We assume that neutrinos travel at speed c . If neutrinos have non-zero mass differences, at least one flavour of neutrino must travel at a lower speed than c and the speed of the two neutrinos must be slightly different. However, for practical purposes, the error introduced by this approximation is negligible.

⁴It is interesting to note that if the baseline satisfies $\Delta m^2 L/E \gg 1$, then the phase will oscillate very rapidly and $\langle \sin^2(\Delta m^2 L/E) \rangle \rightarrow \frac{1}{2}$.

Equations 11.10 and 11.11 show that if at least one pair of neutrinos have different masses and there is some mixing, i.e. $\theta \neq 0$, then transitions between neutrinos of different flavours can occur, violating conservation of lepton flavour number, although not of overall lepton number. The transition probabilities given by eqns 11.10 and 11.11 are simple oscillatory functions of distance and are generally described as ‘oscillation probabilities’; $P(\alpha \rightarrow \beta)$ is referred to as the ν_β *appearance* probability and $P(\alpha \rightarrow \alpha)$ as the ν_α *survival* probability. Before neutrino oscillations were discovered, there was no theoretical guidance as to the value of Δm^2 and it was assumed that any mixing, i.e. θ , would be very small.

Since neutrino masses are now known to be very small, Δm^2 must also be small and L/E must be large for flavour-changing oscillations to be observable. For practical purposes, the dependence of the oscillation probabilities on L/E is usefully expressed as (see Exercise 11.7)

$$P(\alpha \rightarrow \beta) = \sin^2(2\theta) \sin^2\left(\frac{1.27\Delta m^2 L}{E}\right) \quad (11.12)$$

$$P(\alpha \rightarrow \alpha) = 1 - \sin^2(2\theta) \sin^2\left(\frac{1.27\Delta m^2 L}{E}\right) \quad (11.13)$$

where L/E is in km GeV^{-1} or m MeV^{-1} and Δm^2 is in eV^2 . Even with Δm^2 as (unrealistically) large as 1 eV^2 , a detector would have to be 1 km from a source of 1 GeV neutrinos for the phase of the energy-dependent terms in eqns 11.12 or 11.13 to approach $\pi/2$ and for such an experiment to be sensitive to oscillations. The smallness of Δm^2 therefore explains why neutrino oscillations were discovered with neutrinos from natural sources—cosmic rays and the Sun—rather than at particle accelerators.

11.4 Evidence for neutrino oscillations

Detecting neutrinos is difficult because of the very small cross sections involved. The neutrino–proton cross section scales with neutrino energy E_ν as $\sigma \sim G_{\text{F}}^2 m_p E_\nu$, so, for example, $\sigma(\nu_e p) \sim 10^{-41} \text{ m}^2$ for a 10 MeV neutrino.⁵ Therefore, very large active targets are required and/or very intense neutrino sources such as nuclear reactors.

Increased sensitivity to small mass differences can be achieved by using lower-energy neutrinos and making observations at greater distances from the source (see Section 11.3). The actual energies and distances used in real experiments are a compromise between these factors and the requirement to have a measurable reaction rate. Experiments looking for neutrinos produced in the Sun or by cosmic rays in the atmosphere need to be located deep underground to reduce the background from cosmic rays. For the case of low-energy neutrinos studied in solar neutrino oscillations, extreme care must be taken to minimize radioactive backgrounds. For the higher-energy neutrinos produced by accelerators, the detectors can be similar to those used to study neutrino deep inelastic

⁵This is 15 orders of magnitude smaller than the pp total cross section.

Neutrino source	Flavour	E_ν (MeV)	L (m)	$(\Delta m)^2$ (eV ²)
Long-baseline accelerator	μ	10^3 – 10^4	10^6	10^{-2}
Atmospheric	e, μ	10^3	10^7	10^{-3}
$\bar{\nu}_e$, reactor	e	1	10^3	10^{-3}
ν_e , solar	e	1	10^{11}	10^{-11}

Table 11.1 Approximate sensitivities of the different types of neutrino oscillation experiments.

scattering (see Chapter 9). In the case of neutrinos from nuclear reactors or secondary beams from accelerators, if there are detectors at different distances from the source, the oscillations can be measured without needing to know the absolute neutrino flux of the source, thus eliminating a major source of systematic uncertainty. Alternatively, if the distance is fixed, one can study the change in the energy spectrum caused by oscillations. The approximate sensitivity for neutrino mass-squared differences using various sources of neutrinos is given in Table 11.1.

We will first look at the evidence for oscillations from atmospheric neutrinos (Section 11.4.1). The laboratory confirmation of these oscillations will be briefly reviewed in Section 11.4.2. We will then review the evidence for solar neutrino oscillations (Section 11.4.3) and describe the confirmation of these oscillations from experiments using reactor neutrinos (Section 11.4.5). The explanation of the solar neutrino oscillations requires enhanced oscillation probabilities when neutrinos travel through matter (the so-called MSW effect) and a simple explanation will be given in Section 11.4.4. The prospects for studying CP violation in the neutrino sector will be briefly reviewed in Section 11.5.3. Finally, in Section 11.5.4, we will review the most exciting prospect in the area of neutrino oscillations, namely the idea that neutrino oscillations might explain the observed matter–antimatter asymmetry of the Universe.

11.4.1 Atmospheric neutrinos

Neutrinos are produced by cosmic rays (protons and heavier nuclei) colliding with air nuclei in the atmosphere. The flux of high-energy cosmic rays is found to scale with energy E proportionally to $E^{-2.7}$. However, below about 20 GeV, the primary cosmic rays are strongly affected by the Earth’s magnetic field. Even allowing for the linear increase in neutrino cross section with energy, the neutrinos that interact will be predominantly of low energy; the event rate peaks at about 1 GeV. A large background is that from other cosmic rays that interact electromagnetically or strongly and can overwhelm the neutrino signal. This background is greatly suppressed by operating the detectors deep underground.⁶ The primary cosmic rays interact with nitrogen and oxygen

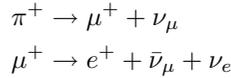
⁶Typical depths are greater than about 1 km.

⁷The atmosphere is approximately 11 nuclear interaction lengths deep, so the probability of a primary cosmic ray not interacting in the atmosphere is negligible.

⁸This detector and others like Soudan were originally designed to search for proton decay, with neutrino interactions being regarded as a source of background.

⁹If a muon stops inside the detector, the subsequent decay will produce a delayed electron that produces a displaced ring. This provides an additional identification power for muons.

in the atmosphere,⁷ producing charged pions, which decay in turn to muon and electron neutrinos according to the decay chain for π^+ (with a similar decay chain for π^-):



Thus, we naively (see Exercise 11.3) expect two muon-type neutrinos for each π^+ but only one electron-type neutrino. The best data come from the Super-Kamiokande experiment, which uses a 50 000 ton water Čerenkov detector to detect the electrons and muons produced by the neutrino interactions in the water.⁸ The Čerenkov light produced by a charged particle travelling with a speed greater than the speed of light in water is emitted in a ring around the direction of the particle's motion (see Chapter 4). Electron neutrinos produce electrons, which in turn produce electromagnetic showers in the water. Therefore, electron neutrinos tend to produce 'fuzzier' rings than muon neutrinos and this can be used to provide a powerful statistical separation between the two flavours of neutrinos. The different responses of the Super-Kamiokande detector to electrons and muons are illustrated in the event displays [129] in Fig. 11.1.⁹ Although there are 20% uncertainties in the absolute neutrino fluxes, some of the uncertainties cancel in the ratio $R = \text{Flux}(\nu_\mu)/\text{Flux}(\nu_e)$. The data are usually presented in terms of the double ratio $R' = R_{\text{experiment}}/R_{\text{predicted}}$, where the prediction assumes no neutrino oscillations. The results from the Super-Kamiokande experiment and other experiments [115] are summarized in Table 11.2. The values of R' are significantly lower than unity, which would be expected in the absence of neutrino oscillations.

Even more direct evidence for neutrino oscillations comes from the zenith-angle distribution. The zenith angle θ_Z is the angle between the vertical and the direction of the incoming neutrino. So values of $\cos\theta_Z$ close to 1 correspond to distances travelled by the neutrinos of ~ 10 km, whereas those with negative values of $\cos\theta_Z$ have travelled for distances of $\sim 10\,000$ km. The distributions from Super-Kamiokande [115] are shown in Fig. 11.2.

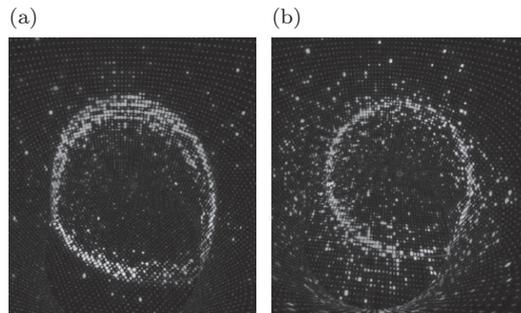


Fig. 11.1 Super-Kamiokande event displays [129] for (a) a muon neutrino event and (b) an electron neutrino event.

Experiments	Double ratio R'	Exposure
Kamiokande-s	$0.60_{-0.05}^{+0.06} \pm 0.05$	7.7
Kamiokande-m	$0.57_{-0.07}^{+0.08} \pm 0.07$	8.2
Soudan-2 (iron)	$0.66 \pm 0.11 \pm 0.06$	4.2
Super-Kamiokande-s	$0.64 \pm 0.02 \pm 0.05$	79
Super-Kamiokande-m	$0.68 \pm 0.03 \pm 0.05$	79

Table 11.2 Measurement of atmospheric neutrino ratio R' . Exposure is in units of kton-years and the suffixes 's' and 'm' stand for sub-GeV and multi-GeV, respectively. The errors quoted are the statistical and systematic errors.

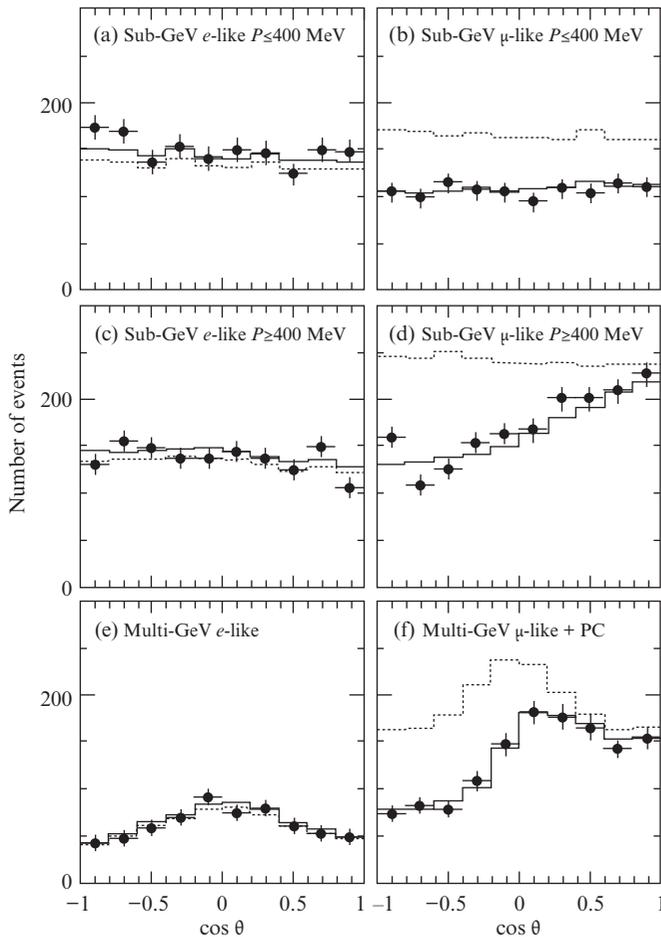


Fig. 11.2 Zenith-angle distributions for the electron-like (a, c, e) and muon-like (b, d, f) events [115]. The events are classified in three energy ranges: less than 400 MeV (a, b), less than 1 GeV (c, d), and greater than 1 GeV (e, f). The dotted histograms show the expectations in the absence of neutrino oscillations and the solid histogram is the result of the fit for $\nu_\mu \rightarrow \nu_\tau$ oscillations.

These distributions show good agreement with the no-oscillation calculations for the electron neutrinos and for the muon neutrinos at large positive values of $\cos\theta_Z$. However, the muon neutrinos show a large deficit at negative values of $\cos\theta_Z$. This is exactly what would be expected for muon neutrino oscillations, because the events at negative values of $\cos\theta_Z$ correspond to neutrinos that have travelled large distances through the Earth ($\sim 10\,000$ km), whereas the events at positive $\cos\theta_Z$ correspond to neutrinos that have only travelled ~ 20 km. The data can be explained by neutrino oscillations of the type $\nu_\mu \rightarrow \nu_\tau$ with $\sin^2 2\theta \sim 1$ and $\Delta m^2 \sim 0.003 \text{ eV}^2$.

11.4.2 Laboratory confirmation of atmospheric neutrino oscillation

We start this section with the description of how a neutrino beam is produced from an accelerator. Then follows a discussion of results from the MINOS experiment.

The NuMI beam at Fermilab

¹⁰We cannot focus neutrinos, but we can produce a somewhat collimated neutrino beam by focusing the charged pions before they decay. The focusing uses toroidal fields that vary as $1/r$ (where r is the radial distance from the beam axis). This ensures that particles emerging at large angles to the beam axis remain in the field region for longer, which in turn ensures that charged particles within a certain angular range emerge parallel to the beam axis. This is achieved using ‘magnetic horns’ that require currents of ~ 100 kA. This is too large for DC currents, so pulsed currents are used (synchronized to the proton bunches).

The creation¹⁰ of a high-intensity, high-energy, laboratory beam of neutrinos (and antineutrinos) starts by extracting the primary 120 GeV proton beam from the Tevatron (2.5×10^{13} protons per pulse) and directing it at a long thin target (see Fig. 11.3). This produces a large flux of charged pions and kaons, with energies in the range 2–60 GeV, which is then focused and collimated, before entering a 675 m evacuated pipe in which the π s and K s decay to $\mu\nu_\mu$ (see Exercise 11.8). The decay pipe is aimed at the MINOS far detector. The next stage is to absorb any remaining hadrons and the large flux of muons. The hadron absorber, consisting mainly of steel, is placed immediately after the decay pipe. Finally, muons must be removed before the NuMI beam passes through the MINOS near detector in a cavern 240 m beyond the hadron absorber. The intervening rock is dolomite and of sufficient density to absorb the muon flux within that distance.

By the time the neutrinos have reached a distant detector, the beam size will have become too large for a detector to contain the beam. Therefore, the detector volume is made as large as can be afforded to

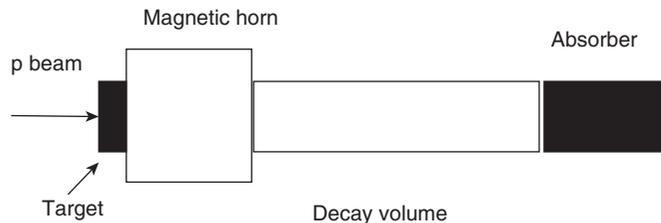


Fig. 11.3 Schematic view of the structure used to create intense neutrino beams.

compensate for the small neutrino cross sections. In practice, the optimal detector shape tends to be approximately cylindrical, with the length along the beam direction being much greater than the width. As very large detectors are required, relatively cheap and simple detector technologies must be employed (see Chapter 4 for a discussion of generic neutrino detectors).

MINOS

For example, the MINOS¹¹ experiment [109] uses magnetized iron as the target, with scintillator layers interspersed with the iron plates. The MINOS detector has two parts: the far detector located at a distance of 735 km from Fermilab and a near detector at Fermilab. The far detector consists of alternating plates of iron and scintillator, with a total mass of 5400 tons.¹² The iron acts as the passive absorber for the calorimeter, but in addition it is magnetized. The scintillator is segmented into narrow strips, which are read out by wavelength-shifting fibres.¹³ These fibres are coupled to photomultiplier tubes via additional clear fibres. The scintillator signals are used for calorimeter measurements as well as for muon tracking. The muons are identified because they penetrate much deeper into the detector. Their momentum can be determined from track curvature in the magnetic field, but if they are contained in the detector, a more precise determination can be made from their range.

The detector is located 716 m underground (in the Soudan mine) in order to minimize backgrounds from cosmic rays. To further reduce backgrounds, veto detectors are located around the main detector. Therefore, the signal for charged-current interactions of ν_μ is the presence of a high-momentum muon. The hadronic energy can be measured by the total energy deposited in the scintillators. A critical feature of the experiment is that there is a near detector close to the origin of the neutrino beam as well as a far detector at a distance of 735 km. The near detector measures the neutrino flux at the origin and can therefore be used to predict the flux and energy spectrum of neutrinos at the far detector in the absence of oscillations. It was the difference between the measured and predicted rates in the far detector that confirmed [10] the presence of muon neutrino oscillations (see Fig. 11.4).

¹¹Main Injector Neutrino Oscillation Search.

¹²Unless otherwise stated, ‘MINOS detector’ refers to the far detector.

¹³The wavelength-shifting fibres are doped with a suitable chemical to convert shorter-wavelength photons to longer, which increases the absorption length. The photons trapped inside the fibre can easily be transported to the photomultipliers using total internal reflection (as in normal optical fibres).

11.4.3 Solar neutrinos

The Sun generates energy by nuclear fusion reactions. The most important reactions are those of the *pp* cycle:

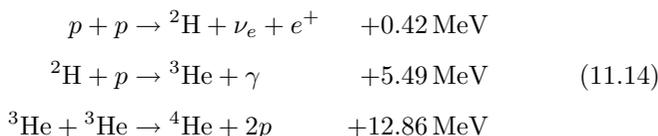
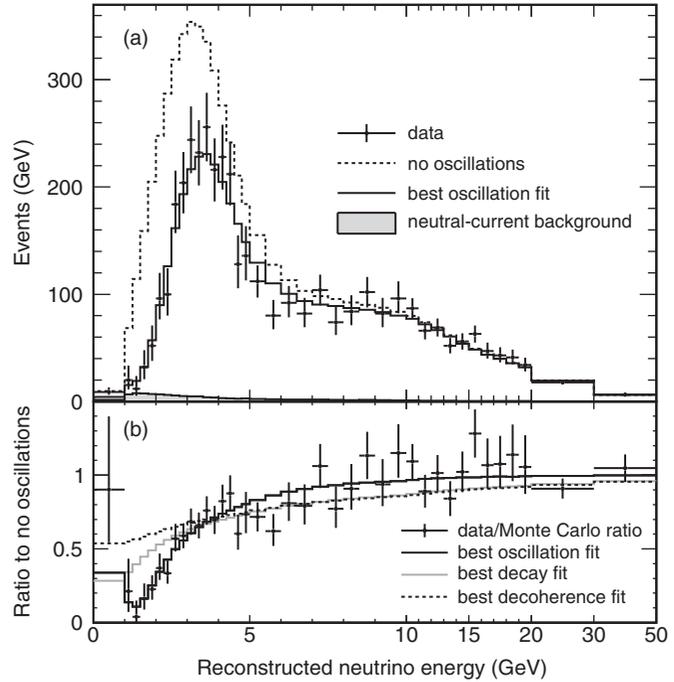
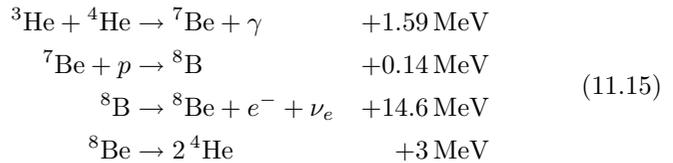


Fig. 11.4 Evidence for ν_μ disappearance from the MINOS experiment. (a) Energy spectrum of the neutrinos as reconstructed in the far detector. The data are compared with the no-oscillation hypothesis and with an oscillation fit. (b) Ratio of the observed spectrum to that expected from the no-oscillation hypothesis [10].



Other branches of the cycle can produce higher-energy neutrinos, which are easier to detect; for example, we can have



where the endpoint of the resulting neutrino spectrum is at 14.6 MeV. The Standard Solar Model (SSM) prediction [42] for the solar neutrino spectrum is shown in Fig. 11.5. Note that the bulk of the spectrum is due to the pp cycle and is therefore at very low energy, which makes the neutrino detection much more difficult.¹⁴

The first technique used to detect solar neutrinos was based on the reaction $\nu_e + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^-$. The chlorine was contained in a tank of 630 tons of C_2Cl_4 ,¹⁵ from which the argon was extracted periodically. The ${}^{37}\text{Ar}$ decays by electron capture (the inverse reaction to that which produced it) and is left in an excited state. The atomic electron capture leaves a ‘hole’ in a low-energy state, and this will be filled by an electron from a higher energy level. The energy released can result in an outer electron being ejected from the atom (the ‘Auger’ effect). The resulting Auger electrons were detected in a proportional counter (see Chapter 4). This experiment was started over 40 years ago by Ray Davies and was the first to detect solar neutrinos and show that the rate was lower than

¹⁴The low-energy neutrinos from the pp cycle have been measured using Ga detectors; however, the main evidence for neutrino oscillations relies on measurements of the higher-energy neutrinos.

¹⁵ C_2Cl_4 (perchloroethylene) is often used in dry cleaning.

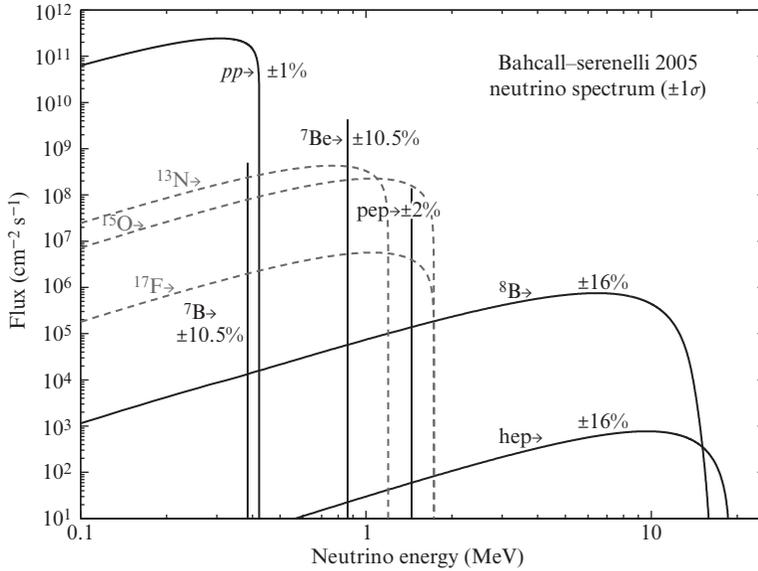


Fig. 11.5 Predicted neutrino energy spectrum in the SSM [115].

expected. Davies shared the 2002 Physics Nobel Prize with Masatoshi Koshiha (Kamiokande) and Riccardo Giacconi.¹⁶

Several experiments have seen a clear deficit of electron neutrinos compared with the predictions of the SSM. It is very difficult to reconcile these data with modifications of the SSM, whereas they could all be explained by electron neutrinos oscillating into other flavours [115]. However, the most model-independent demonstration that the neutrino deficit is due to oscillations rather than problems with the SSM comes from the SNO experiment [12]. The SNO experiment used 1000 tons of very pure D₂O viewed by 10 000 large photomultipliers. The primary reactions are the charged-current (CC) interactions of the electron neutrinos,



the neutral-current (NC) interactions of all neutrino flavours,



and the elastic scattering (ES) reaction



The Feynman diagrams for the three processes are shown in Fig. 11.6. The CC interactions are sensitive only to electron neutrinos because the thresholds of the equivalent reactions with other neutrino flavours are too high. The NC interactions are equally sensitive to all neutrino flavours. All three neutrino flavours contribute to the ES reactions, but the cross section is much larger (~ 6 times) for electron neutrinos. Therefore, by measuring the rate for CC, NC, and ES interactions, i.e. enough information to deduce the total $\nu_e + \nu_\mu + \nu_\tau$ rate, one can look

¹⁶The former director of the Space Telescope Science Institute.

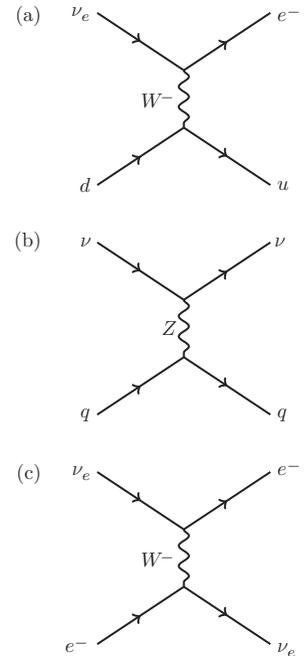


Fig. 11.6 Feynman diagrams at the quark level for charged-current (a), neutral-current (b), and elastic scattering (c) reactions.

for evidence of neutrino flavour transitions, independent of the SSM. Unambiguous evidence for neutrino flavour transitions is then provided by the measured flux of all neutrino flavours (as determined by the NC interactions) and the flux of electron neutrinos as determined from the CC interactions. The signals for the CC and ES interactions come from the produced electrons generating electromagnetic showers, which then lead to Čerenkov radiation, which is detected in the photomultipliers. The NC interactions are more difficult to detect.

In the first phase of SNO, neutrinos were detected by neutron capture on deuterium, which results in 6.25 MeV photons, which produce electromagnetic showers that then lead to Čerenkov radiation. In order to increase the sensitivity of the experiment to NC interactions, in the second phase of SNO operation (SNO_s), 2 tons of NaCl were added to the D₂O because this enabled neutron capture on ³⁵Cl. The ³⁵Cl has a high absorption cross section for low-energy neutrons, and neutron capture leads to photons with an energy distribution peaked around 8 MeV. The measured fluxes from SNO (in units of 10⁶ cm⁻² s⁻¹) are [12]

$$\begin{aligned}\phi_{CC} &= 1.68 \pm 0.06 \text{ (stat.) } {}^{+0.08}_{-0.09} \text{ (syst.)} \\ \phi_{ES} &= 2.35 \pm 0.22 \text{ (stat.) } \pm 0.15 \text{ (syst.)} \\ \phi_{NC} &= 4.94 \pm 0.21 \text{ (stat.) } {}^{+0.38}_{-0.24} \text{ (syst.)}\end{aligned}\quad (11.19)$$

This combination of measurements gives the determination of the flux of muon and tau neutrinos (in units of 10⁶ cm⁻² s⁻¹) as

$$\phi(\nu_\mu) + \phi(\nu_\tau) = 3.26 \pm 0.25 \text{ (stat.) } {}^{+0.40}_{-0.35} \text{ (syst.)}\quad (11.20)$$

This observation of a non-zero flux of ν_μ or ν_τ from the Sun therefore provides model-independent evidence of solar neutrino oscillations. The SNO results are combined with those from Super-Kamiokande [115] and are shown in Fig. 11.7. The data are clearly inconsistent with the no-oscillation hypothesis ($\phi(\nu_\mu) + \phi(\nu_\tau) = 0$). All the data are consistent with each other and with the predictions of the SSM. Finally, the combined results from all phases of the SNO experiment can

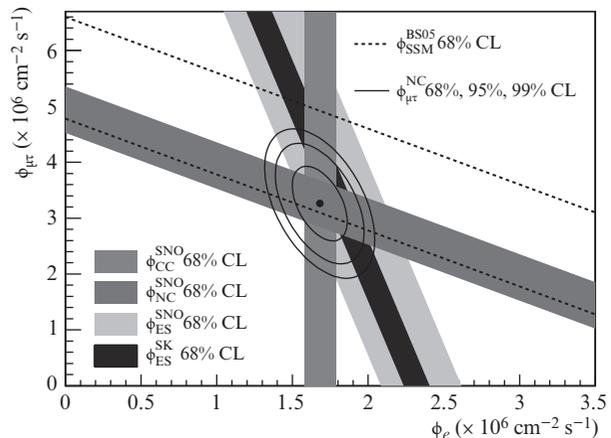


Fig. 11.7 Fluxes of ⁸B solar neutrinos, deduced from the SNO charged-current, neutral-current, and elastic scattering results from the salt-phase measurement and the results from Super-Kamiokande. The vertical axis is the flux of muon and tau neutrinos $\phi_{\mu\tau}$ and the horizontal axis is the flux of electron neutrinos ϕ_e . The expectations from the SSM are shown by dashed lines [115].

also be used to measure the total flux of ${}^8\text{B}$ neutrinos¹⁷ from the Sun $\phi_{\text{NC}}^{\text{SNO}} = 5.25 \pm 0.16$ (stat.) ${}_{-0.13}^{+0.11}$ (syst.), which is in good agreement with the SSM prediction of $\phi_{\text{SSM}} = 5.94{}_{-0.81}^{+1.01}$. This confirms that the Sun is producing energy from nuclear fusion at the rate expected in the SSM.

Although the SNO results confirm the hypothesis that neutrino oscillations occur, in the context of the vacuum oscillation model we have considered it is difficult to understand why the ratio of electron neutrino to all flavours of neutrinos ν_e/ν_x is $< \frac{1}{2}$. If the oscillations are occurring rapidly compared with the Sun–Earth distance, then the minimum value of the ratio would be $\frac{1}{2}$. This statement is strictly only true in a two-flavour model, but this is a good approximation for solar neutrinos. Before considering the probable explanation for this, we first need to consider other solar neutrino experiments that are sensitive to lower-energy neutrinos. Most of the total flux of neutrinos comes from the first reaction in the pp chain (see eqn. 11.14), which generates neutrinos with a continuous spectrum with an endpoint at 0.42 MeV (see Fig. 11.5). The flux of these low-energy neutrinos is very tightly constrained by the observed solar luminosity, unlike the flux of higher-energy neutrinos, which are more sensitive to details of the solar model (in particular the core temperature). Therefore, the original motivation for efforts to detect these low-energy neutrinos was to see if the solar neutrino deficit was due to the solar model or to neutrino oscillations. However, from our perspective, these data are particularly interesting when compared with the higher-energy data, because they reveal a significant energy dependence in the oscillation probability.

The detection of very low-energy neutrinos ($E_\nu < 0.42$ MeV) is challenging. The radioactive backgrounds in water Čerenkov detectors are too large to allow a sufficiently low threshold to be set. Such low-energy neutrinos have been studied by radiochemical experiments using the reaction $\nu_e + {}^{71}\text{Ga} \rightarrow {}^{71}\text{Ge} + e^-$, which has a threshold energy of 233 keV. Slightly over half of the interactions are expected to come from the low-energy pp neutrinos. In the GALLEX experiment, the produced Ge was periodically extracted (about every 10 days) chemically and the number of ${}^{71}\text{Ge}$ atoms was detected using electron capture (essentially the inverse of the reaction that created the ${}^{71}\text{Ge}$). The electron capture leaves a hole in the K or L shell, which is filled by electrons from higher energy states, resulting in X-ray emission. The X-rays are detected in proportional counters filled with xenon gas (xenon is a good absorber for X-rays because of its large Z value).¹⁸ The results from the different gallium experiments are summarized in [115] and show a suppression compared with the prediction of the SSM of a factor of about $\frac{1}{2}$. Comparing these results for low-energy neutrinos with those from SNO and Super-Kamiokande, which are sensitive to higher-energy neutrinos, we can see that the neutrino oscillation probability has a significant energy dependence. These observations are hard to reconcile with vacuum oscillations, so we need to consider the effect of matter on neutrino oscillations, which we will do in the next section.

¹⁷The threshold used by SNO to reduce the backgrounds meant that the detector was mainly sensitive to ${}^8\text{B}$ neutrinos.

¹⁸The target contains 101 tons of GaCl_3 and a typical 10-day run produces about ten ${}^{71}\text{Ge}$ atoms. Therefore, the detector has to be deep underground to be shielded from cosmic rays, and extreme care must be taken to minimize radioactive backgrounds.

11.4.4 MSW effect

There is very strong evidence for neutrino oscillations, as we have seen in Sections 11.4.1–11.4.3 (see also Section 11.4.5). The atmospheric neutrino data can be explained by the phenomenology of vacuum oscillations (see Section 11.3) with a close to maximal mixing angle between ν_μ and ν_τ . This implies that we can consider the electron neutrino mixing in the simple two-component picture (the ν_e mixes with a linear superposition of ν_μ and ν_τ). Therefore, the probability of a ν_e created in the Sun remaining as a ν_e is given by

$$P_{\nu_e \rightarrow \nu_e}(t) = 1 - \sin^2 2\theta \sin^2 \left(\frac{\Delta m^2 t}{4E} \right) \quad (11.21)$$

For $\Delta m^2 > 10^{-4} \text{ eV}^2$, the phase factor oscillates so rapidly that it will average to a value of $\frac{1}{2}$, and therefore the minimum value of $P_{\nu_e \rightarrow \nu_e}(t)$ would be $\frac{1}{2}$. However, the SNO and other data clearly indicate a significantly lower value (see Section 11.4.3).

This apparent paradox can be understood if we allow for the effects of matter on the propagation of neutrinos through the Sun, the so-called Mikheyev–Smirnov–Wolfenstein (MSW) effect¹⁹ [108, 135]. The effective Hamiltonian differs from its vacuum counterpart by the addition of weak interactions. All flavours of neutrinos can have neutral-current interactions, but only (low-energy) electron neutrinos can undergo charged-current interactions. From the perspective of oscillations, we are only interested in the differences between electron neutrinos and muon or tau neutrinos, and therefore we do not need to consider the neutral-current interactions. The probability of incoherent scattering in the Sun is negligible and in any case could not contribute to interference effects. We are therefore only interested in the charged-current coherent forward scattering, which gives a contribution to the Hamiltonian for electron neutrinos²⁰ of

$$H(r) = \sqrt{2} G_F N_e(r) \quad (11.22)$$

where G_F is the Fermi coupling constant and $N_e(r)$ is the electron volume number density at a radius r in the Sun.²¹

To understand the MSW effect, we start by considering the Schrödinger equation for two neutrino flavours propagating in vacuum. For oscillations, we are only interested in the terms in the Hamiltonian that are different for electron neutrinos compared with other flavours of neutrinos. We can write this part of the Hamiltonian as

$$H_V = \frac{\Delta m^2}{4p} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} \quad (11.23)$$

and the time-independent Schrödinger equation as

$$H_V \begin{pmatrix} \nu_e \\ \nu_x \end{pmatrix} = E \begin{pmatrix} \nu_e \\ \nu_x \end{pmatrix} \quad (11.24)$$

¹⁹Here we give a simplified description of the MSW effect in order to illustrate how it can resolve this apparent paradox.

²⁰We are working in the flavour basis for the neutrino eigenstates.

²¹The effect on neutrino propagation is similar to a change in refractive index in the optical case. This then changes the propagation speed for electron neutrinos and hence the oscillation rate.

where ν_x represents the non-electron neutrino flavours (i.e. those that do not interact with electrons by charged-current interactions). We can then show that the difference in the energies of the two eigenstates is given by $\Delta E = \Delta m^2/2p$ (see Exercise 11.5). The effect of the charged-current coherent forward scattering is to change the effective potential for electron neutrinos by $V_e = \sqrt{2} G_F N_e$, where N_e is the electron number density. We can evaluate the effect on the Hamiltonian using $E^2 - p^2 = m^2$ and assuming that neutrinos are ultrarelativistic and $V_e \ll E$:

$$m_\nu^2 = (E + V_e)^2 - p^2 \approx m^2 + 2EV_e \quad (11.25)$$

Therefore, the change in m^2 for the electron neutrino is given by

$$\Delta m_{\nu_e}^2 = 2\sqrt{2} G_F N_e E \quad (11.26)$$

Again assuming that the neutrinos are ultrarelativistic, we can write the contribution from matter to the Hamiltonian as

$$\Delta H_M = \sqrt{2} G_F N_e \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (11.27)$$

As usual, we can drop any term proportional to the unit matrix from the point of view of oscillations. It is therefore convenient to rewrite this as a term proportional to the unit matrix and the matrix relevant for oscillations:

$$\Delta H_M = \sqrt{2} G_F N_e / 2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (11.28)$$

We can then combine the vacuum term, eqn 11.23, with the matter term, eqn 11.28, to obtain the Hamiltonian in the presence of matter:

$$H_M = \frac{\Delta m^2}{4p} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} + \frac{\sqrt{2} G_F N_e}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (11.29)$$

It is conventional to define $A = 2\sqrt{2} G_F N_e p / \Delta m^2$, simplifying eqn 11.29 to give

$$H_M = \frac{\Delta m^2}{4p} \begin{pmatrix} -\cos 2\theta + A & \sin 2\theta \\ \sin 2\theta & \cos 2\theta - A \end{pmatrix} \quad (11.30)$$

We can then define an effective mixing angle in the presence of matter as θ_m and define the total effective Hamiltonian in the presence of matter as

$$H_M = \frac{\Delta m^2}{4p} \begin{pmatrix} -\cos 2\theta_m & \sin 2\theta_m \\ \sin 2\theta_m & \cos 2\theta_m \end{pmatrix} \quad (11.31)$$

Comparing eqn 11.30 with eqn 11.31, we can relate the mixing angle in the presence of matter to that in vacuum:

$$\tan 2\theta_m = \frac{\tan 2\theta}{1 - A \sec 2\theta} \quad (11.32)$$

This clearly has a resonant condition if $A > 0$ and $A = \cos 2\theta$, or, substituting for the definition of A , we find the electron neutrino resonant energy at which the mixing becomes maximal

$$E_{\text{Res}} = \frac{\Delta m^2 \cos 2\theta}{2\sqrt{2} G_{\text{F}} N_e} \quad (11.33)$$

We can now use this MSW formalism to explain how the measured suppression of electron neutrinos can be energy-dependent and how the suppression of high-energy electron neutrinos can be larger than a factor of two, which would be the maximal value (assuming that the electron neutrinos make many oscillations between the Sun and the Earth).

²²Solar neutrinos are made by nuclear fusion in the core of the sun. As the core of the sun is much smaller than the full volume, they will be created near the centre.

Electron neutrinos with energy $E > 2 \text{ MeV}$ at the centre of the Sun²² will have a higher energy than E_{Res} (see eqn 11.33 and Exercise 11.6). Since the density of the Sun increases towards the centre (owing to gravity), the electron density decreases smoothly as the neutrinos leave; therefore, these electron neutrinos will hit the resonance condition. As the density changes slowly with radius, this will happen ‘adiabatically’ and the neutrinos that will propagate out to the surface of the Sun will be the heavier-mass eigenstate ν_2 :

$$\nu_2 = \nu_e \sin \theta + \nu_\mu \cos \theta \quad (11.34)$$

where θ is the vacuum mixing angle. Since these states are now eigenstates of H_V , they will simply propagate to the Earth with no further oscillations. From eqn 11.34, we can then easily see that the ν_e survival probability is given by (eqn 11.11)

$$P(\nu_e \rightarrow \nu_e) = \sin^2 \theta \quad (11.35)$$

and can therefore be less than $\frac{1}{2}$. Electron neutrinos with lower energies will not see this resonance and will effectively propagate as vacuum states. As these neutrinos are oscillating rapidly compared with the transit time from the Sun to the Earth, the average phase factor will be $\frac{1}{2}$, so the survival probability will be given by

$$P(\nu_e \rightarrow \nu_e) = 1 - \frac{1}{2} \sin^2 2\theta \quad (11.36)$$

The combined results of the solar neutrino experiments can be fitted in terms of the MSW effect (see Section 11.4.4), with $\Delta m^2 \sim 5 \times 10^{-5} \text{ eV}^2$ and $\tan^2 \theta \sim 0.45$ [115].

11.4.5 Confirmation of solar neutrino oscillations

Confirmation that neutrino oscillation was the correct solution to the solar neutrino puzzle came from the KamLAND experiment. The detector is based deep underground in the Kamiokande site in Japan. It detects electron antineutrinos from a large number of Japanese nuclear reactors. The flux-weighted average distance from the sources to

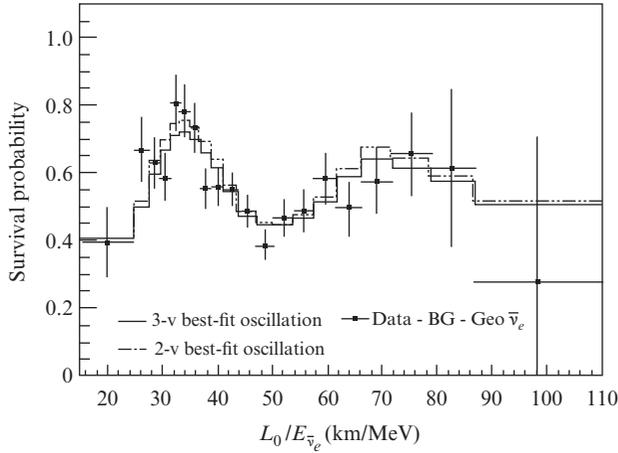


Fig. 11.8 Ratio of the measured neutrino rate divided by that expected in the no-oscillation model, as a function of $L_0/E_{\bar{\nu}_e}$. The lines are best fits for neutrino oscillation models [77].

the detector is about $L_0 = 180$ km. This long baseline provides sensitivity to small values of the mass splitting. The primary reaction is $\bar{\nu}_e + p \rightarrow e^+ n$. The detector consists of 1000 tons of liquid scintillator, surrounded by photomultipliers. The advantage of liquid scintillator over water (Čerenkov) is that it produces a larger signal, thus enabling it to achieve lower energy thresholds. This is important because the neutrino energies are below 8 MeV. However, extreme care must still be taken to minimize radioactive isotopes in the detector. The primary signal comes from the e^+ but there is also a delayed signal from γ rays after the neutrons are captured on protons. This helps reduce the background and enables the threshold to be lowered to 2.6 MeV [77].²³

The KamLAND experiment measured the $\bar{\nu}_e$ spectrum and, by dividing by the expected flux from the no-oscillation model, it was possible to measure the survival probability as a function of L_0/E (where E is the neutrino energy). The results in Fig. 11.8 show clear evidence for neutrino oscillations and, even more interestingly, these data show the characteristic maxima and minima expected from oscillations.

From a combined fit to the solar and KamLAND data, the solution to the solar neutrino problem requires the MSW effect.

²³There is also a background from ‘geothermal’ neutrinos, i.e. neutrinos from the decays of radioactive isotopes in the Earth. In future, the study of geothermal neutrinos might become interesting for geophysics research for distinguishing between different models of the Earth’s core.

11.5 Three (or more)-flavour oscillations

11.5.1 Generalized oscillation probabilities

Expressions for the oscillation probabilities for the general case of N flavours and N massive neutrinos can be derived in the same way as for two flavours. The N flavour eigenstates $|\nu_\alpha\rangle$ are expressed as linear combinations of the N mass eigenstates $|\nu_k\rangle$:

$$|\nu_\alpha\rangle = \sum_k U_{\alpha k} |\nu_k\rangle$$

where the coefficients $U_{\alpha k}$ are the elements of an $N \times N$ PMNS matrix. In general, the $U_{\alpha k}$ are complex, and unitarity requires that

$$\begin{aligned}\sum_{\alpha} U_{\alpha k} U_{\alpha j}^* &= \delta_{jk} \\ \sum_k U_{\alpha k} U_{\beta k}^* &= \delta_{\alpha\beta}\end{aligned}$$

If a neutrino of flavour α is created at $t = 0$, the initial state is

$$|\psi(0)\rangle = \sum_k U_{\alpha k} |\nu_k\rangle$$

At some later time t , the phases of the mass eigenstates have evolved and

$$|\psi(t)\rangle = \sum_k U_{\alpha k} e^{-iE_k t} |\nu_k\rangle$$

where E_k is the energy of the ν_k . The mass eigenstates can be expressed in terms of the flavour states using²⁴

$$|\nu_k\rangle = \sum_{\alpha} U_{\alpha k}^* |\nu_{\alpha}\rangle$$

so

$$|\psi(t)\rangle = \sum_{\beta} \left(\sum_k U_{\alpha k} e^{-iE_k t} U_{\beta k}^* \right) |\nu_{\beta}\rangle$$

The amplitude for observing a ν_{β} at time t is

$$\langle \nu_{\beta} | \psi(t) \rangle = \sum_k U_{\alpha k} U_{\beta k}^* e^{-iE_k t}$$

Then, as for the two-flavour case, taking $E_k \approx p + m_k^2/2p$ and $p = E$,

$$\langle \nu_{\beta} | \psi(t) \rangle = \sum_k (U_{\alpha k} U_{\beta k}^* e^{-im_k^2 t/2E}) e^{-iEt}$$

and the probability of observing a neutrino of flavour β at a distance L from the source is

$$\begin{aligned}P(\nu_{\alpha} \rightarrow \nu_{\beta}) &= |\langle \nu_{\beta} | \psi(t = L) \rangle|^2 \\ &= \sum_{k,j} U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j} \exp\left(-i \frac{\Delta m_{kj}^2 L}{2E}\right)\end{aligned}\quad (11.37)$$

where $\Delta m_{kj}^2 \equiv m_k^2 - m_j^2$.

²⁴As the matrix U is unitary, the inverse is $U^{-1} = U^{\dagger}$.

The neutrino oscillation probability, eqn. 11.37, can also be written as

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} - 4 \sum_{k>j} \operatorname{Re}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j}) \sin^2 \left(\frac{\Delta m_{kj}^2 L}{4E} \right) \\ &\quad + 2 \sum_{k>j} \operatorname{Im}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j}) \sin \left(\frac{\Delta m_{kj}^2 L}{2E} \right) \end{aligned} \quad (11.38)$$

For antineutrinos, the elements of the PMNS matrix $U_{\alpha k}$ must be replaced by their complex conjugates $U_{\alpha k}^*$ and

$$|\bar{\nu}_\alpha\rangle = \sum_k U_{\alpha k}^* |\bar{\nu}_k\rangle$$

The antineutrino oscillation probabilities become

$$\begin{aligned} P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) &= \delta_{\alpha\beta} - 4 \sum_{k>j} \operatorname{Re}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j}) \sin^2 \left(\frac{\Delta m_{kj}^2 L}{4E} \right) \\ &\quad - 2 \sum_{k>j} \operatorname{Im}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j}) \sin \left(\frac{\Delta m_{kj}^2 L}{2E} \right) \end{aligned} \quad (11.39)$$

The third terms of eqns 11.38 and 11.39 have different signs, with the important consequence that for the same L/E the neutrino and antineutrino oscillation probabilities are different. Since neutrinos and antineutrinos are CP conjugates, this immediately suggests the possibility of CP violation and, by the CPT theorem, T violation. The CP asymmetry would be

$$\begin{aligned} A_{\alpha\beta}^{CP}(L, E) &= P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) \\ &= 4 \sum_{k>j} \operatorname{Im}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j}) \sin \left(\frac{\Delta m_{kj}^2 L}{2E} \right) \end{aligned} \quad (11.40)$$

CP violation would be manifested as a difference between the $\nu_\alpha \rightarrow \nu_\beta$ and $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ oscillation probabilities, and T violation by a difference in the time-reversed probabilities $P(\nu_\alpha \rightarrow \nu_\beta)$ and $P(\nu_\beta \rightarrow \nu_\alpha)$. For CP or T violation to occur, the imaginary parts of eqn 11.40 must not vanish; in other words, the PMNS matrix elements $U_{\alpha k}$ must be complex.²⁵ CP violation would not occur for only two neutrino flavours since, just as in the case of two quark generations, the mixing can be described by a single angle.

The transition $\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha$ is the CPT conjugate of $\nu_\alpha \rightarrow \nu_\beta$ and, by the CPT theorem, $P(\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha) = P(\nu_\alpha \rightarrow \nu_\beta)$. Therefore, the CP -conjugate survival probabilities $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\alpha)$ and $P(\nu_\alpha \rightarrow \nu_\alpha)$ must be equal, and CP violation can only be observed by comparing appearance (i.e. flavour-changing) probabilities.

The rather formidable expressions for the oscillation probabilities in eqns 11.38 and 11.39 simplify in certain cases. For example, when one of

²⁵The quantity $\operatorname{Im}(U_{\alpha k} U_{\beta k}^* U_{\alpha j}^* U_{\beta j})$ is directly analogous to the invariant area J of the CKM unitarity triangles, eqn 10.28 in Section 10.6, which describes CP violation in the quark sector.

the Δm_{kj}^2 is much greater than all other Δm^2 's ('one-mass-scale dominance'), they reduce to quasi-two-flavour expressions described by a single effective mixing angle.

11.5.2 Three-flavour oscillations

There is currently no convincing evidence for more than three neutrino flavours. The mixing of the known neutrinos can therefore be described in terms of three mixing angles θ_{12} , θ_{13} , and θ_{23} , a single phase δ , and the three mass (squared) differences, Δm_{12}^2 , Δm_{23}^2 , and Δm_{13}^2 between three mass eigenstates ν_1 , ν_2 , and ν_3 . It is known that $\Delta m_{13}^2 \approx \Delta m_{23}^2 \gg \Delta m_{12}^2$ since solar and atmospheric neutrino oscillations are described by two very different mass differences. From the solar neutrino data allowing for the MSW effect (see Section 11.4.4), we can determine that $\Delta m_{12}^2 > 0$. However, we do not currently know the hierarchy of the masses m_2 and m_3 ; i.e. is $\Delta m_{23}^2 < 0$?

The three-flavour PMNS matrix (eqn 11.1)

$$\mathbf{U} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \quad (11.41)$$

can be expressed in terms of the mixing angles as

$$\mathbf{U} = \begin{pmatrix} c_{12}c_{13} & c_{13}s_{12} & s_{13}e^{-i\delta} \\ -c_{23}s_{12} - c_{12}s_{13}s_{23}e^{i\delta} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta} & c_{13}s_{23} \\ s_{23}s_{12} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - c_{23}s_{12}s_{13}e^{i\delta} & c_{13}c_{23} \end{pmatrix} \quad (11.42)$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The mixing angles θ_{ij} can be thought of as representing rotations around a third, k , axis. For example, θ_{13} represents a rotation around the '2' axis.²⁶ The matrix can also be written as follows to make the three separate rotations more apparent:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (11.43)$$

It can be seen from eqn 11.42 that the phase δ must be non-zero for CP violation to occur. It can also be shown that all the angles, including θ_{13} , must be non-zero for CP violation to occur.

The three-flavour oscillation probabilities for neutrinos and antineutrinos can be written in terms of the mixing angles by substituting the $U_{\alpha i}$ of eqn 11.42 into eqn 11.38 or 11.39, respectively, for $N = 3$. Since the resulting expressions are rather long, they will not be given here. An important result, however, is that the dependence of the oscillation probabilities on L/E falls into three different regimes. If $\Delta m_{13}^2 L/E \ll 1$, no

²⁶This representation of the PMNS matrix is identical with the representation of the CKM matrix of quark mixing discussed in Section 10.6. Of course the angles have entirely different physical meanings.

flavour oscillations will occur. In the second regime, where $\Delta m_{23}^2 L/E \approx 1$ but $\Delta m_{12}^2 L/E \ll 1$, the oscillation probability $P(\nu_\alpha \rightarrow \nu_\beta)$ depends mainly on the four elements of \mathbf{U} that couple ν_α and ν_β to ν_2 and ν_3 , i.e. $U_{\alpha 2}$, $U_{\beta 2}$, $U_{\alpha 3}$, and $U_{\beta 3}$. Finally, if $\Delta m_{12}^2 L/E \geq 1$, the oscillation probabilities depend on all the $U_{\alpha i}$. For example, if $U_{e3} = 0$ (as is approximately true), ν_e does not couple to ν_3 and $\nu_e \leftrightarrow \nu_\mu$, and $\nu_e \leftrightarrow \nu_\tau$ oscillations will not occur unless $\Delta m_{12}^2 L/E \geq 1$, although $\nu_\mu \rightarrow \nu_\tau$ oscillations can take place at smaller L/E as long as $\Delta m_{23}^2 L/E \approx 1$. The two regimes $\Delta m_{12}^2 L/E \ll 1 \leq \Delta m_{23}^2 L/E$ and $1 \leq \Delta m_{12}^2 L/E$ are often referred to as the ‘atmospheric’ and ‘solar’ oscillation regimes, respectively.

When the results of all neutrino oscillation experiments and measurements are combined, it is found that the mixing is described by

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} 0.82 & 0.56 & \sim 0.15 \\ 0.31-0.43 & 0.51-0.59 & 0.75 \\ 0.37-0.47 & 0.59-0.66 & 0.66 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (11.44)$$

and $\Delta m_{23}^2 \approx \Delta m_{13}^2 = 2.3 \times 10^{-3} \text{ eV}^2$ and $\Delta m_{12}^2 = 7.5 \times 10^{-5} \text{ eV}^2$. At present there is no experimental evidence that the phase δ is non-zero, and the values in eqn 11.44 are the magnitudes of $U_{\alpha i}$. It can be seen that since $|U_{e3}| \equiv \sin \theta_{13} \approx 0.15$, ν_e is coupled only weakly to ν_3 ; the ν_3 ‘content’ of ν_e is $|U_{e3}^2| \approx 0.02$, i.e. only 2%. From the discussion in the preceding paragraph, this means that ν_e barely participates in oscillations involving ν_2 and ν_3 only, i.e. where $\Delta m_{23}^2 L/E \approx 1$ but $\Delta m_{12}^2 L/E \ll 1$, and justifies the analysis of the results of atmospheric neutrino and long-baseline accelerator experiments in terms of two-flavour oscillations characterized by Δm_{23}^2 and θ_{23} . It also means that the disappearance of solar ν_e can be described in terms of two flavours characterized by Δm_{12}^2 and θ_{12} .²⁷ The oscillation regimes are illustrated in Fig. 11.9.

In stark contrast to the CKM matrix, which is approximately diagonal, all the elements of \mathbf{U} , with the exception of U_{e3} (equivalently θ_{13}), are large. Their approximate equality suggests that CP violation by neutrinos could be large for the same reason that CP violation by mesons depends on the size of the CKM matrix elements via the angles of the unitarity triangles as discussed in Section 10.6.

²⁷This can relatively easily be seen by setting $s_{13} = 0$ and $c_{13} = 1$ in eqn 11.42 and deriving the ν_e survival probability $P(\nu_e \rightarrow \nu_e)$.

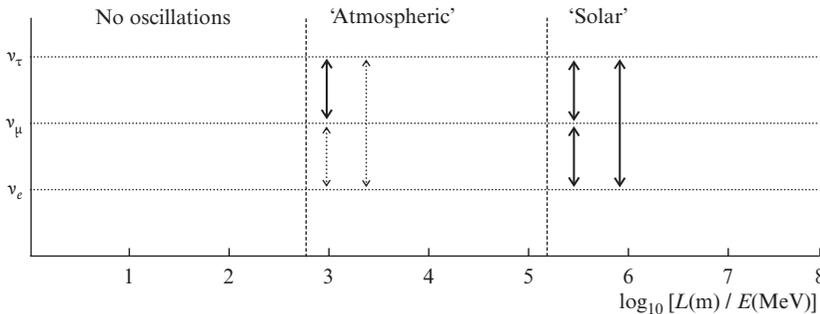


Fig. 11.9 The behaviour of flavour-changing neutrino oscillations as a function of L/E . In the atmospheric regime, the transitions are predominantly $\nu_\mu \leftrightarrow \nu_\tau$; in the solar regime, all flavours participate.

11.5.3 Measurement of the mixing angle θ_{13}

Very convincing evidence for a non-zero value of θ_{13} was shown by the Daya Bay experiment. This used an array of six identical detectors to measure the $\bar{\nu}_e$ flux around six nuclear reactors. The detectors used liquid scintillator viewed by photomultipliers. The reaction is the same as that used by KamLAND (see Section 11.4.5) and is also a disappearance experiment, but over baselines of ~ 1 km, rather than the ~ 100 km studied by KamLAND, which is why there is sensitivity to the mixing angle θ_{13} . There was a separate layer of purified water outside the scintillator to act as a veto detector (for example to reject incoming cosmic rays). The detectors nearer the reactors measured the flux, and this was used to predict the flux at the further away detectors. The ratio of measured to predicted flux at the further away detectors was significantly less than 1, which is evidence for a non-zero value of θ_{13} . This result was confirmed by RENO (another reactor experiment). Further confirmation of the non-zero value of θ_{13} was provided by the T2K experiment. This used a very intense 30 GeV proton beam at the JPARC accelerator to send a neutrino beam to the Super-Kamiokande detector at a distance of 295 km. In order to obtain the maximal oscillation probability for this baseline and the expected value of Δm_{13}^2 , the optimal neutrino energy was $E_\nu \sim 1$ GeV. This was achieved by having the beam line at an angle of 2.5° with respect to the line from the accelerator to the detector. Clear evidence for ν_e appearance in a ν_μ beam is shown by the spectrum of identified ν_e interactions (see Fig. 11.10) [8].

A global fit to these and other data gave a value of $\sin^2 2\theta_{13} = 0.096 \pm 0.013$ [115]. A necessary but not sufficient condition for CP violation in the neutrino sector is that θ_{13} is non-zero. Therefore, the observation of this large value for θ_{13} opens the way to the study of CP violation in the neutrino sector, which would be very interesting for the reasons discussed in Section 11.5.4. This would require measuring different oscillation rates for neutrinos and antineutrinos. Several ideas

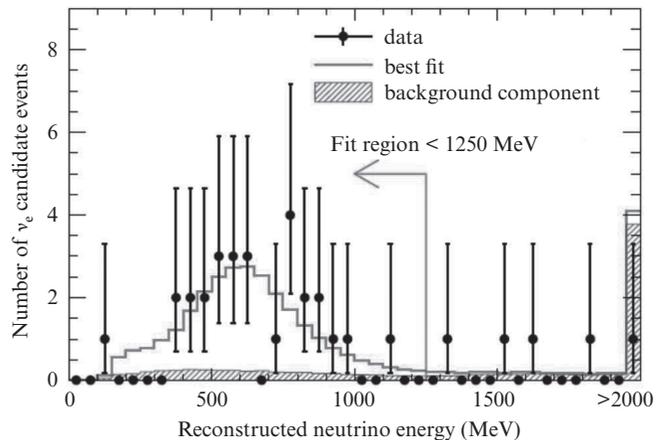


Fig. 11.10 The energy spectrum of detected electron neutrinos from the T2K experiment with the Super-Kamiokande detector. The data are significantly above the estimated background and is in good agreement with the expectations from neutrino oscillations [8]. The solid curve is the result of a fit to the oscillation model.

for new very long-baseline experiments are being considered. They are very challenging because potential small differences between neutrino and antineutrino oscillation probabilities need to be measured, which will require very intense neutrino beams as well as very large detectors.

11.5.4 Matter–antimatter asymmetry

The Universe appears to be dominated by matter compared with antimatter. If there were stars or galaxies made of antimatter, we would expect characteristic γ rays from matter–antimatter annihilation. There is no evidence for this from astrophysics. The primary cosmic-ray flux contains only a small component of antiparticles such as positrons, and these can be explained by secondary production processes. As well as explaining this effect, we also need to understand the baryon-to-photon (n_B/n_γ) ratio. In the early Universe, with temperatures satisfying $k_B T \gg 2m_p$, we can have the reactions $p\bar{p} \leftrightarrow \gamma\gamma$, which would have been in thermodynamic equilibrium, and the value of n_B/n_γ would be determined by the Boltzmann and statistical factors. However, as the Universe cooled, the interaction rate became lower than the inverse of the expansion time, so the baryons would have been ‘frozen out’. The Standard Model prediction²⁸ for $n_B/n_\gamma \sim 10^{-19}$ is much lower than the measured value $n_B/n_\gamma \sim 10^{-10}$.

²⁸Using the CKM mechanism.

If the Universe started off with matter–antimatter symmetry, then we must satisfy the three Sakharov conditions [124] in order to have the currently observed matter-dominated Universe:²⁹

²⁹The explanation of the matter–antimatter asymmetry can also resolve the problem with the ratio n_B/n_γ .

- (1) There must be baryon number violation.
- (2) There must be C -symmetry and CP -symmetry violation.
- (3) Interactions must occur at an epoch in which the Universe was not in thermodynamic equilibrium.

The first condition is obvious, but the second is not so self-evident. If C symmetry held, there would be an equal rate for the production of baryons and antibaryons. Similarly, if CP symmetry held, there would be an equal rate for the production of ‘left-handed’ baryons and ‘right-handed’ antibaryons and an equal rate for the production of ‘left-handed’ antibaryons and ‘right-handed’ baryons. Therefore, the second condition is required to produce more baryons than antibaryons. If the Universe were in thermodynamic equilibrium, CPT symmetry would ensure that there was an equal rate for reactions creating baryons and antibaryons.

The SM does have CP violation in the quark sector, but it turns out that the off-diagonal CKM elements are so small that this cannot provide sufficient CP violation to explain the observed matter–antimatter asymmetry and n_B/n_γ . One possible explanation would be to invoke a Grand Unified Theory (GUT), which would naturally generate baryon-number-violating interactions since quarks and leptons are in the same multiplets. However, these GUT models tend to predict a rate of proton decay that is incompatible with measurements.

There are many different theoretical attempts to explain the observed asymmetry, although here we consider only one model. An attractive possibility is that the CP violation arises in the neutrino sector. We already know that the magnitudes of the off-diagonal elements of the neutrino mixing (PMNS) matrix are relatively large. If the one complex phase, δ , in the PMNS matrix turned out to be large, this would predict relatively strong CP violation in the neutrino sector. Now that we know that neutrinos have non-zero masses, they could be Dirac or Majorana particles (see Chapter 6).³⁰ We will consider a model in which the neutrinos are Majorana particles. One of the unexplained features of the SM is that the neutrino masses are so much smaller than those of the charged leptons and the quarks. One attempt to explain this is motivated by GUTs in which the neutrino mass matrix is given by

$$\mathbf{M}_\nu = \begin{pmatrix} 0 & M \\ M & B \end{pmatrix} \quad (11.45)$$

The value of M is of the order of the electroweak scale, but the value of the parameter B is of the order of the GUT scale, so $B \gg M$. The masses of the neutrinos are then given by the eigenvalues of the mass matrix (eqn 11.45):

$$\lambda_\pm = \frac{1}{2}(B \pm \sqrt{B^2 + 4M^2}) \quad (11.46)$$

In the regime $B \gg M$, this gives $\lambda_+ = B$ and $\lambda_- = -M^2/B$.³¹ Therefore, larger values of B increase (decrease) the mass of the heavier (lighter) eigenstate—hence this is called the ‘see-saw’ mechanism. The lighter mass corresponds to the observed left-handed (LH) neutrino states and the heavier mass state would be a right-handed (RH) neutrino.

In the early Universe with temperatures $k_B T > m(\nu_R)$, these LH and RH neutrinos would have been in thermodynamic equilibrium and there could have been no asymmetry. As the Universe cooled to lower temperatures, it would no longer be in thermodynamic equilibrium (thus satisfying the third Sakharov condition) and the abundance of the heavy RH neutrinos would have been ‘frozen-out’. They would have then decayed to the light neutrino states in a CP -violating way to create a lepton number asymmetry. The lepton asymmetry can be converted to a baryon asymmetry in the SM. There is no perturbative SM process in the SM that can violate $B + L$ (where B is the baryon number and L the lepton number), but this can happen non-perturbatively. The $SU(2)$ vacuum has an infinite number of minima, each with a different value of $B - L$. At low temperatures, the probability of transition from one vacuum state to another is suppressed by a factor of $e^{-M_W/T}$, where M_W is the mass of the W boson. However, at very high temperatures $k_B T > M_W$, the Universe can easily hop over the barrier from one vacuum state to another, thus allowing violation of $B - L$. This then allows the lepton number asymmetry created by the decays of the ν_R

³⁰In Chapter 8, we looked at the search for neutrinoless double β decay, which if observed would prove that neutrinos are Majorana particles.

³¹We require the masses to be positive, so clearly the parameter M must be imaginary.

to be converted to a baryon number (and lepton number) asymmetry. As electric and colour charge are protected by gauge symmetries, the Universe remains neutral for electric and colour charge.

Chapter summary

- Very convincing evidence for neutrino oscillations has been observed in the study of atmospheric neutrinos and confirmed by laboratory experiments.
- Similarly convincing evidence for neutrino oscillations has been found with solar neutrinos and confirmed by a reactor antineutrino experiment.
- In a simple extension of the minimal Standard Model, neutrinos can have non-zero masses. According to quantum mechanics, it then becomes possible for one flavour of neutrino to oscillate into another flavour.
- The discovery of a non-zero mixing angle θ_{13} opens the way for the search for CP violation in the neutrino sector, which might explain the matter–antimatter asymmetry in the Universe.

Further reading

- Perkins, D. (2009). *Particle Astrophysics* (2nd edn). Oxford University Press. This is a very clear and accessible textbook on the many areas in which astrophysical measurements can yield interesting constraints on particle physics.
- Akhmedov, E. Kh. and Smirnov, A. Yu. (2009). Paradoxes of neutrino oscillations. *Phys. Atom. Nuclei*, **72**, 1363. This gives a very thorough discussion of a correct description of neutrino oscillations.

Exercises

- (11.1) Show that we obtain identical formulae for neutrino oscillation probabilities if we assume that the two mass eigenstates have the same energy but different momenta, as opposed to the assumption of the same momentum but different energy used to derive eqn 11.5.
- (11.2) (a) Consider the decay $\pi \rightarrow \mu\nu_i$, where ν_i is a mass eigenstate with mass m_i . Assume there are two neutrino mass eigenstates, one with $m_1 = 0$ and with a splitting in mass given by $\Delta m_{12}^2 = 10^{-4} \text{ eV}^2$. Determine the momenta and energies of the two neutrino mass eigenstates (in the pion CMS), and hence show that neither the momenta nor the energies are the same, and determine the difference in momenta of the two flavours of neutrinos.
- (b) If the decay is localized in some linear region of length l , use the Heisenberg uncertainty principle to determine the uncertainty σ_p in the momentum of the neutrino. For what value of l would σ_p be equal to the difference in momenta between ν_1 and ν_2 states. Comment on your results.
- (11.3) The result that we get two atmospheric muon neutrinos for each electron neutrino, in a given

neutrino energy interval, is not obvious given the steeply changing cosmic-ray spectrum. Starting from a 2 GeV pion, compute roughly the average energy of the ν and μ into which it decays. Then, again using a rough calculation, divide the energy of the μ equally among the decay products to show that all the neutrino energies are roughly equal, so the naive result works. This is a lucky coincidence; repeat this starting with a 2 GeV kaon to show that the sharing is not equal.

- (11.4) SN1987A was a supernova explosion that occurred at a distance of approximately 170 000 light years from Earth. The burst of neutrinos preceded the observation of the supernova with optical telescopes (this was expected because it takes time for the shock wave to reach the surface). The neutrino pulse lasted for about 10 s and the neutrino energy range was about 10–20 MeV. How could such neutrinos be produced in the supernova and how could they be detected? Use these data to estimate an approximate upper limit for the neutrino mass.
- (11.5) Find the eigenvalues of the Hamiltonian given in eqn 11.23. Hence show that the difference in energies of the two eigenstates is given by $\Delta E = \Delta m^2/2p$, as expected for ultrarelativistic neutrinos.

Assuming that neutrinos are ultrarelativistic, derive eqn 11.23. *Hint:* You can drop any term in

the Hamiltonian that is proportional to the unit matrix, since it does not affect oscillations.

- (11.6) Estimate the resonant neutrino energy, assuming that the electron density at the centre of the Sun is $N(r=0) \approx 3 \times 10^{31} \text{ m}^{-3}$.
- (11.7) Explain the factor of 1.27 in eqn 11.12. If $\Delta m^2 = 0.003 \text{ eV}^2$ and $L = 700 \text{ km}$, determine the neutrino energy required to maximize the effect of oscillations.
- (11.8) If we wish to make a neutrino beam from the decays of π^+ with a neutrino energy $\sim 3 \text{ GeV}$, make an approximate estimate of how long a decay path is required.
- (11.9) Calculate the neutrino threshold energy for the three reactions given by eqns 11.16–11.18, assuming that all neutrino masses are negligible. How can the reactions be identified experimentally and why is it necessary to require a minimum recoil electron energy of about 5 MeV in practice?
- (11.10) For elastic νe scattering at a fixed incident neutrino energy E_ν , find a relation between E_e and θ , the energy and angle relative to the incident ν direction of the recoil electron. Assume that the electron is initially at rest. Show that your relation implies that the observed distribution of recoil electrons is strongly peaked in the incident ν direction. Why is this important for studies of solar neutrinos?

The Higgs boson

12

The Higgs mechanism is a part of the standard model that we postponed from the discussion of electroweak unification in Chapter 7. Until recently, it was one possible theoretical solution to an otherwise problematic part of the GSW theory that we have highlighted in earlier chapters—that we want the W and Z to be massive to make a short-range force, but we want them to be massless so that gauge invariance works and the theory is renormalizable (Section 7.4.4). The Higgs mechanism is as old as the rest of the GSW theory and predicts that there should be a boson, the Higgs, but does not predict its mass.

Recently, strong evidence for a boson that fits the description has been found in several different decay channels at both of the big experiments at the LHC. In this chapter, we will first discuss the theory behind the Higgs boson and then take a look at the experimental evidence for its existence.

We start at a ‘middle difficulty’ level, i.e. try to explain all the key ideas without too much mathematics.¹ Then, having set the scene, we give a more sophisticated mathematical picture of the Higgs mechanism in Section 12.4.

The discussion of the Higgs mechanism starts with all the particles being massless. We then postulate a new field with which all particles can interact. The interaction of a particle with this field produces an effect that makes the particle appear to have mass; i.e. the field produces an effect when the particle moves and this manifests itself as *inertia*. To begin with, we have a free hand to choose whatever form of new field we like, but it must respect both Lorentz invariance and local gauge invariance.

The type of field that we need has already been introduced in Section 7.2.6. Now there are certain properties that the new field must possess: it must respect Lorentz covariance, it must be colourless and uncharged, and it must have spin zero.² The field is related to the weak interaction, to give masses to the W and Z , so we will allow it to have both weak isospin and hypercharge (recall that these are related to the electric-charge equivalents for the W^+ , W^0 , W^- and for the B^0 , respectively; see Section 7.4.1). Finally, it must be locally gauge-invariant, because we want it to make the whole theory gauge-invariant.

12.1 Local gauge invariance	334
12.2 Spontaneous symmetry breaking	334
12.3 Higgs mechanism—the simplified story	336
12.4 Lagrangians	337
12.5 Higgs mechanism—more mathematical	339
12.6 Higgs discovery	342
12.7 Coupling to fermions	349
12.8 Determination of the spin and parity of the new boson	351
12.9 Outlook	353
Chapter summary	353
Further reading	354
Exercises	354

¹Specifically, as we made clear in earlier chapters, we are not using the language of relativistic quantum field theory (RQFT). How this affects our treatment of the Higgs mechanism will be covered briefly in Section 12.4.

²We must be relativistically consistent and also not allow the field to interact via any of the known forces. Otherwise, when our massless particle (e.g. an electron) interacts with it, thus gaining mass, it could also exchange spin or charge with the field. This would give the impression that the electron spin or charge could change spontaneously, which would not agree with experiment.

12.1 Local gauge invariance

We give a brief reminder of what gauge invariance entails. It is an important property that fields can, but do not have to, possess. Those that are locally gauge-invariant are, by 't Hooft's theorem, renormalizable, which is important for the reasons described in Section 7.4.4. We have already considered gauge symmetry in the familiar context of classical electric and magnetic fields (see Chapter 6). We showed that the classical observable fields \mathbf{B} and \mathbf{E} are unchanged by the simultaneous changes to \mathbf{A} and ϕ as follows:

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\Lambda, \quad \phi \rightarrow \phi' = \phi - \frac{\partial\Lambda}{\partial t} \quad (12.1)$$

where Λ is a scalar. Λ can be a function of \mathbf{x} and t and, when it is, we call it a *local* gauge transformation. In the context of relativistic quantum mechanics, we showed that the Dirac equation for a particle of charge q in an electromagnetic field (\mathbf{A}, ϕ) is gauge-invariant if the Dirac wavefunction ψ is also transformed:

$$\psi \rightarrow \psi' = \psi e^{iq\Lambda} \quad (12.2)$$

The effect of the simultaneous gauge transformations on the \mathbf{A} , ϕ , and ψ fields on the Dirac equation

$$\left(i\frac{\partial}{\partial t} - q\phi \right) \psi = \boldsymbol{\alpha} \cdot (-i\nabla - q\mathbf{A})\psi + \beta m\psi \quad (12.3)$$

leave the physics unchanged (see Exercise 12.1). From demanding that Λ can be different for different space-time coordinates (local gauge symmetry), we can determine the form of the electromagnetic interaction.

12.2 Spontaneous symmetry breaking

We want to arrange for there to be a Higgs field so that a massless particle can feel an interaction with it and act as if it has a mass. All fields we have encountered so far have been 'in the vacuum', i.e. they are normally switched off (apart from quantum fluctuations) if nothing is happening, so the ground state is zero. Our new field is different. It must be on all the time to generate particle masses.

The simplest type of field is a **scalar field** ϕ , which is a single real number that can vary as a function of \mathbf{x} and t . To make it gauge-invariant, we will have to make it a complex number $\phi = \phi_1 + i\phi_2$. We need to understand what is involved with making the field be on all the time. We start with a field that is real, so we set $\phi_2 = 0$.

Familiar classical fields have energy associated with them when they are on; for example, there is energy associated with a charged capacitor (which is stored in its electric field) that is not present when the capacitor is discharged. This is represented in Fig. 12.1, which shows the potential energy associated with a field as a function of the value of the field.

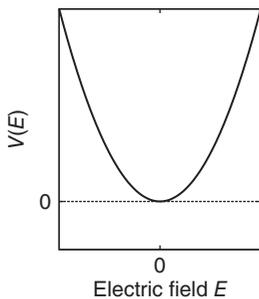


Fig. 12.1 Potential energy associated with a field such as the electric field in a capacitor.

It has a minimum, the value of the field when there is nothing there, which is usually taken to set the zero of the potential. Examples of how the potential energy V of a field with this property could vary with the value of ϕ are $V \propto \phi^2$, $V \propto \phi^4$, and $V \propto \cosh \phi - 1$. There are many others.

Thinking pictorially, we can imagine a ball rolling around in the bottom of the potential well shown in Fig. 12.1. If it has no energy, it will sit in the bottom at $\phi = 0$. If it has some energy, it can oscillate by rolling around in the bottom of the well.

Most of the time when we are using quantum mechanics, we calculate using Feynman diagrams or by applying perturbation theory, which is an expansion in terms of a ‘small’ parameter, for QED the fine-structure constant α . The order of perturbation theory (or Feynman diagrams) required will depend on the accuracy required. This is analogous to using a Taylor series expansion to calculate a mathematical function; however, going beyond the first order in RQF raises some difficult mathematical issues. To make perturbation theory work, we need a region in which ϕ is slowly varying. For the field theory case in which we are interested, this corresponds to expanding about the minimum of the potential. For example, for a potential with a minimum at $\phi = 0$ as shown in Fig. 12.2, we can consider expansions about $\phi = 0$.

Now consider an example of a field that does not have its potential energy minimum at $\phi = 0$, as shown in Fig. 12.3. The potential in this example varies as

$$V(\phi) = \frac{1}{4}\lambda^2\phi^4 - \frac{1}{2}\mu^2\phi^2 \quad (12.4)$$

where λ and μ are both real constants. This has two minima, which are at $\phi = \pm\mu/\lambda$. The vacuum state of the field is now not the one with ‘no-field’ ($\phi = 0$) but the one with either $\phi = \pm\mu/\lambda$. This means that we need to consider quanta of the field as excitations with respect to this non-zero field value, and the vacuum acquires a non-zero ‘vacuum expectation value’ (often shortened to ‘vev’ in RQF texts). This is the key idea of spontaneous symmetry breaking. The fundamental theory respects a symmetry but the symmetry is ‘hidden’ (or broken) in the vacuum state. If such a field were to exist in nature, we would have to do our Feynman diagram expansions about one of the minima, a procedure that we will explore in this chapter. These ideas are worked through for this potential in Exercise 12.6. In terms of a classical analogy and thinking pictorially with a ball in a potential well, the ball can now rattle around in the bottom of either minimum.

However, as noted above, the real scalar field considered so far does not work for the Higgs mechanism since it is not possible to make it locally gauge-invariant. So we now consider a complex scalar field $\phi = \phi_1 + i\phi_2$ in which ϕ_2 does not have to be zero. We will use the potential

$$V = \frac{1}{4}\lambda^2(\phi^*\phi)^2 - \frac{1}{2}\mu^2(\phi^*\phi) \quad (12.5)$$

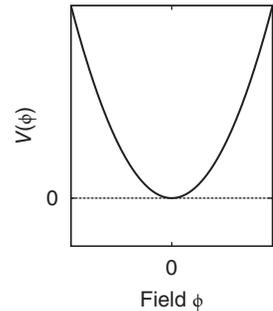


Fig. 12.2 Potential energy function with a minimum at $\phi = 0$.

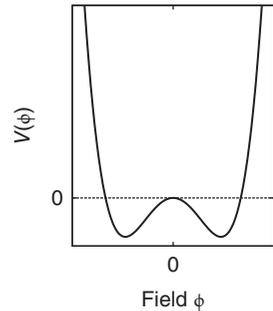


Fig. 12.3 Potential energy function $V(\phi)$ for a field with a minimum at non-zero values of ϕ .

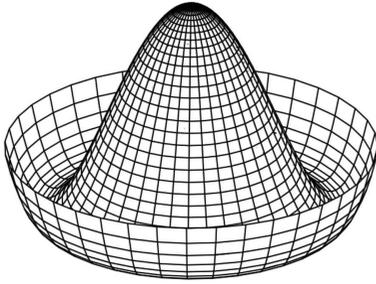


Fig. 12.4 ‘Mexican hat’ potential for the Higgs field. The vertical axis represents the potential and the horizontal axes represent the real and imaginary parts of the complex Higgs field. From Millard, Rupert. Higgs Mexican hat potential. http://commons.wikimedia.org/wiki/File:Mexican_hat_potential_polar.svg.

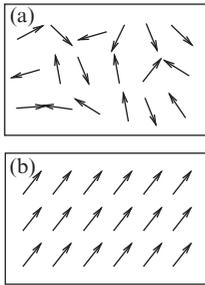


Fig. 12.5 Spontaneous symmetry breaking in an isotropic ferromagnet: (a) symmetry above the Curie temperature; (b) hidden symmetry below the Curie temperature.

³This discussion is greatly simplified. In a real ferromagnet below the Curie temperature, the spins are aligned within small domains but the orientation of the spins in different domains is random.

which is a logical extension of the potential we have used up to now when we change from real ϕ to complex ϕ . If we substitute $\phi = \phi_1 + i\phi_2$ into this potential, we get

$$V = \frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) \quad (12.6)$$

This function is shown in Fig. 12.4. The potential has a likeness to a Mexican hat; the minimum has become a ring when $\phi_1^2 + \phi_2^2 = (\mu/\lambda)^2$ and the phase of the complex number is arbitrary. If we picture a ball rolling around in this potential, we can see two degrees of freedom. One of these has the ball rolling up and down the sides in the same way as for the ball in Fig. 12.3; i.e. $\phi_1^2 + \phi_2^2$ oscillates and the phase of the complex number stays fixed. The other way, which is new, is for the ball to roll around the bottom; i.e. the value of $\phi_1^2 + \phi_2^2$ stays fixed, but the phase of the complex number changes. This second degree of freedom requires no energy for it to occur. In quantum field theory, it corresponds to the existence of a boson called a ‘Goldstone boson’, which has no mass. No such boson exists experimentally, but we will address this shortly.

The concept we have been exploring here is called spontaneous symmetry breaking. The potential is symmetric (i.e. the minimum can be anywhere around the rim of the Mexican hat), but in nature this symmetry is hidden, i.e. the field has to pick one particular value. Another example of spontaneous symmetry breaking in nature comes from the spin alignment behaviour in an isotropic ferromagnet. In Fig. 12.5(a), above the Curie temperature, the spins are randomly aligned. The situation is isotropic, i.e. rotationally symmetric. In Fig. 12.5(b), below the Curie temperature, the spins align (an electrostatic effect due to quantum-mechanical wavefunctions of the electrons).³ The symmetry is now **hidden**. It is spontaneously broken. The symmetry still exists, because any direction for the spin alignment could equally be chosen. This is analogous to the Higgs field that is inserted into the standard model.

12.3 Higgs mechanism—the simplified story

In Section 12.4.2, we will give a mathematical description of how the above concepts fit together to make the Higgs mechanism. In this section, we give a simple picture of the ideas underlying the Higgs mechanism. We start with the equation for a spin-0 particle that obeys the potential we have just described, $V = \frac{1}{4}\lambda^2\phi^4 - \frac{1}{2}\mu^2\phi^2$, which can behave as if it were ‘on’ all the time. This contains a degree of freedom corresponding to the ball rolling round the bottom of the potential, which requires no energy and causes the theory to predict a massless boson, the ‘Goldstone boson’.

We then add a ‘vector’ (spin-1) field that is locally gauge-invariant. This field has to be massless (in order to respect the gauge invariance) before we consider the effect of the non-zero expectation value of the

scalar (Higgs) field. After spontaneous symmetry breaking, the vector field interacts with the scalar field and acquires a mass because of the non-zero vacuum expectation value of the scalar field. We then extend this in a way that allows four fields rather than just one to be added; these four fields fit the properties of the W^\pm , W^0 , and B^0 (from the first three steps of the electroweak unification procedure from Section 7.4.1). By making a few subtle rearrangements of the equation, we are able to organise the terms so that

- the Goldstone boson disappears;
- the W^\pm and the Z^0 appear to have a mass;
- the photon remains massless;
- $M_W/M_Z = \cos \theta_W$.

So, almost miraculously, the new term in the potential makes everything come out exactly as we need in order to agree with experiment.

Furthermore, we can repeat the procedure with a fermion, by adding it in a similar way and insisting on local gauge invariance, with the cancellation of terms occurring between the Higgs field and the fermion, and this generates the mass term for the fermion. We can keep doing this with all the fermions in the Standard Model (SM) to get masses for each of them. The theory does not predict the mass of each fermion, but it does predict that the couplings $g_f \bar{f} H$ are proportional to the masses of the fermions. The theory also does not specify the mass of the Higgs boson.

12.4 Lagrangians

Before moving on to applying the Higgs mechanism in more mathematical detail, we shall make a short detour. Lagrangians are a very powerful alternative to Newton's laws in classical mechanics.⁴ The Lagrangian approach also works in quantum mechanics and it provides the mathematical framework for RQF. These ideas were touched on briefly at the start of Chapter 6. While the Dirac equation allows only a single particle or a system of a fixed number of particles to be considered, RQF allows the number of particles to change. RQF can cope with creation and annihilation of fermion pairs, for example, and with having several different fields within one equation at the same time.

⁴We have deliberately avoided using Lagrangian mechanics in this book because of the mathematical overheads involved. However, this approach is particularly useful for discussing the Higgs boson, so we give a flavour of it here that may be helpful when consulting more advanced texts.

12.4.1 Lagrangians in classical mechanics

Lagrangian mechanics follows a specific recipe,⁵ which has three steps. The first step is to pick the correct number of independent variables for the system. For example, a pendulum that moves in one dimension has one independent variable, θ . A pendulum that can swing in any horizontal direction has two variables; we will choose the angle to the vertical θ and the azimuthal angle in the horizontal plane ϕ . The second

⁵Why it works is explained in many textbooks on classical mechanics.

⁶By definition, $\mathcal{L} = T - V$.

step is to write down an expression for the Lagrangian \mathcal{L} by working out the total kinetic energy of the system T and the total potential energy of the system V .⁶ This must be in terms of the independent variables and their first derivatives with respect to time (so for the example of the two-dimensional pendulum, these are $\theta, \dot{\theta}, \phi, \dot{\phi}$). For the two-dimensional pendulum with a mass m attached to a string of length l , the Lagrangian is

$$\mathcal{L} = \frac{1}{2}m(l\dot{\theta})^2 + \frac{1}{2}m(l\dot{\phi}\sin\theta)^2 - (1 - \cos\theta)mgl \quad (12.7)$$

where g is the acceleration due to gravity.

The third step is to use the Euler–Lagrange equations with \mathcal{L} . There is one Euler–Lagrange equation for each independent variable and each gives an equation of motion (which are usually coupled). We could have tried to start directly with these equations of motion, but one of the beauties of the Lagrangian method is to be able to manipulate everything in one equation before having to deal with simultaneous equations. The Euler–Lagrange equations written for the two variables in our two-dimensional pendulum problem are (see Exercise 12.2)

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}} \right) = \frac{\partial \mathcal{L}}{\partial \theta}, \quad \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\phi}} \right) = \frac{\partial \mathcal{L}}{\partial \phi} \quad (12.8)$$

The partial derivative with respect to θ is taken holding all of the other variables and their derivatives ($\dot{\theta}, \phi,$ and $\dot{\phi}$) constant, and similarly for the other partial derivatives. For the pendulum, the resulting equations of motion are

$$l\ddot{\theta} + g\sin\theta = 0, \quad 2\dot{\theta}\dot{\phi}\sin\theta\cos\theta + \ddot{\phi}\sin^2\theta = 0 \quad (12.9)$$

A general feature of Lagrangians is that if an independent variable does not appear explicitly in the Lagrangian, but only its time derivative (in our example, ϕ does not appear explicitly), then the right-hand side of the corresponding Euler–Lagrange equation is zero. Such a coordinate is called cyclic and there is a conserved quantity associated with it; in our example, this is $\partial\mathcal{L}/\partial\dot{\phi} = ml^2\sin^2\theta\dot{\phi} = L_\phi$. This result is immediately apparent from the ϕ Euler–Lagrange equation.⁷

Another general feature is the ability to move terms between being considered as kinetic or potential energy, giving what is referred to as an effective potential V_{eff} . In the example, we can change the Lagrangian by inserting L_ϕ , which is constant:

$$\mathcal{L} = \frac{1}{2}m(l\dot{\theta})^2 + \frac{L_\phi^2}{2ml^2\sin^2\theta} - (1 - \cos\theta)mgl \quad (12.10)$$

We can now pretend that the second term is part of the potential energy (even though it started as part of the kinetic energy) and we have an equation involving only one degree of freedom moving in a fictitious potential describing our two-dimensional pendulum. This feature of swapping terms around within the Lagrangian carries over to the use of Lagrangians in RQF and is very helpful for understanding the Higgs mechanism.

⁷In this case, we can identify L_ϕ as the component of angular momentum about the z axis. In the Lagrangian approach, we can see that conservation laws (e.g. angular momentum) are a result of the invariance of the Lagrangian with respect to a conjugate variable, in this case ϕ .

12.4.2 Lagrangians in quantum mechanics

We shall not dwell too much on Lagrangians in quantum mechanics. The terms in the Lagrangian still correspond to what we loosely call kinetic and potential energy. Just as it is possible to have several objects in one classical Lagrangian, it is possible to express a system with many fields as a single Lagrangian. For each field in the Lagrangian, there is an analogue to the classical Euler–Lagrange equations,⁸ and different terms in the Lagrangian produce recognizable expressions on application of the Euler–Lagrange equations. For example, terms in the Lagrangian from a spin-0 field give the Klein–Gordon equation, while terms from a spin- $\frac{1}{2}$ field give the Dirac equation. This also applies to spin-1 fields, and in the case of the photon (a spin-1 field with zero mass), one can use the Lagrangian approach to generate Maxwell’s equations.

⁸These have more parts to them than the classical ones, however, because of the need to maintain Lorentz invariance.

12.5 Higgs mechanism—more mathematical

We are going to repeat the discussion above, but now following the terms that appear in the equations more closely. We will use Lagrangians since the formalism is easier to follow and this will provide an introduction to the detailed explanations in more advanced textbooks. We will build up the concepts that go into the Higgs mechanism with four examples. The particular potentials used here are examples only, and there are a number of different fields that can be made to work theoretically, but using these examples gives a good insight into how the mechanism works in principle.

Example 1

We start with the Lagrangian for a spin-0 particle with no potential energy term:

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi)}_{\text{K.E. term}} - \underbrace{\frac{1}{2}m^2\phi^2}_{\text{mass term}} \quad (12.11)$$

Applying the Euler–Lagrange equation to this (see Exercise 12.4) gives the Klein–Gordon equation for a spin-0 particle, $\partial_\mu\partial^\mu\phi + m^2\phi = 0$. The second term in the Lagrangian gives rise to the mass term. For it to be a mass term, it must be proportional to the square of the field and it must be negative.

Example 2

We now modify this by setting $m = 0$ (so the mass term disappears) but we add a potential $V = \frac{1}{2}\mu^2\phi^2$ (remember that $\mathcal{L} = T - V$):

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi)}_{\text{K.E. term}} - \underbrace{\frac{1}{2}\mu^2\phi^2}_{\text{P.E.}} \quad (12.12)$$

Notice how the second term looks just like a mass term. If the particle is in the field that has produced this type of potential, it will behave as if it has a mass. Now we change to a more sophisticated potential of the type discussed earlier, $V = \frac{1}{4}\lambda^2\phi^4 - \frac{1}{2}\mu^2\phi^2$:

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi)}_{\text{K.E. term}} - \underbrace{\frac{1}{4}\lambda^2\phi^4}_{\text{interaction}} + \underbrace{\frac{1}{2}\mu^2\phi^2}_{\text{not a mass term}} \quad (12.13)$$

Terms that are third or fourth powers of a field (or combination of fields) represent interactions in RQF. So the ϕ^4 term above represents an interaction. The final term in this Lagrangian is a problem—we do not know how to interpret it. It is not a mass term, because it has the wrong sign.

However, if we change our zero point in the field, what happens? To do this, we need to re-express the field ϕ in terms of a new field ρ that is zero at the bottom of one of the two minima. Let $\phi = \rho \pm \mu/\lambda$, where the \pm reflects the fact that there are two minima that can be used. We get (see Exercise 12.6):

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\rho)(\partial^\mu\rho)}_{\text{K.E.}} - \underbrace{\mu^2\rho^2}_{\text{mass}} \pm \underbrace{\mu\lambda\rho^3 - \frac{1}{4}\lambda^2\rho^4}_{\text{interactions}} \quad (12.14)$$

⁹We have dropped a constant term because it disappears when the Euler–Lagrange equations are applied, so we can ignore it.

Now we have a mass term!⁹

Example 3

We need to make ϕ complex to proceed with local gauge invariance. Let $\phi = \phi_1 + i\phi_2$, where ϕ_1 and ϕ_2 are real, and consider the potential $V = \frac{1}{4}\lambda^2(\phi^*\phi)^2 - \frac{1}{2}\mu^2\phi^*\phi = \frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2)$. The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi_1)(\partial^\mu\phi_1) + \frac{1}{2}(\partial_\mu\phi_2)(\partial^\mu\phi_2) - \frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 + \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) \quad (12.15)$$

The first two terms are kinetic energy terms. Next, as in Example 2, we expand around a minimum. In this case, we choose $\phi_1 = \rho + \mu/\lambda$ and $\phi_2 = \rho'$. This is where the minimum is on the positive real axis. We could do it about any point, and would get the same result (but perhaps after quite a lot of algebra and some redefinition of the fields). The Lagrangian is now

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\rho)(\partial^\mu\rho) - \mu^2\rho^2}_{\rho \text{ scalar, } m=\sqrt{2}\mu} + \underbrace{\frac{1}{2}(\partial_\mu\rho')(\partial^\mu\rho')}_{\rho' \text{ scalar, no mass}} - \underbrace{\mu\lambda(\rho^3 + \rho\rho'^2) - \frac{1}{4}\lambda^2(\rho^4 + \rho'^4 + 2\rho^2\rho'^2)}_{\text{interactions}} \quad (12.16)$$

The ρ field becomes a massive field with mass $\sqrt{2}\mu$ and the ρ' is a massless field (the Goldstone boson we discussed earlier). There are several interaction terms and we have omitted a constant term.

Example 4

Next, we apply local gauge invariance, by adding a massless vector field, i.e. a field with spin 1 like a photon field. There is a prescription—it makes the terms we are familiar with in the local gauge invariance discussion cancel properly. We start from Example 3:¹⁰

$$\mathcal{L} = \overbrace{\frac{1}{2}[(\partial_\mu - iqA_\mu)(\phi_1 - i\phi_2)][(\partial^\mu + iqA^\mu)(\phi_1 + i\phi_2)]}^{\text{K.E. term modified for local gauge invariance}} - \underbrace{\frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 + \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2)}_{-V} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{K.E. for vector field}} \quad (12.17)$$

¹⁰This equation is eqn 13.108 from Burcham and Jobes and eqn 10.129 from Griffiths—see Further Reading.

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ (see Exercise 12.5). We now expand, as is becoming familiar, about a minimum in the potential: $\phi_1 = \rho + \mu/\lambda$ and $\phi_2 = \rho'$. The result is

$$\mathcal{L} = \overbrace{\frac{1}{2}(\partial_\mu\rho)(\partial^\mu\rho) - \mu^2\rho^2}^{\rho \text{ scalar, } m=\sqrt{2}\mu} + \overbrace{\frac{1}{2}(\partial_\mu\rho')(\partial^\mu\rho')}^{\rho' \text{ scalar, no mass}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{vector K.E.}} + \underbrace{\frac{1}{2}q^2\frac{\mu^2}{\lambda^2}A_\mu A^\mu}_{\text{vector mass}} - \underbrace{q\frac{\mu}{\lambda}A_\mu\partial^\mu\rho'}_{\text{problem}} \quad (12.18)$$

+ interaction terms

This is now locally gauge-invariant. The spin-1 field has acquired a mass (the second term on the second line). However, there remains a problem, namely the third term on the second line, which looks like an interaction that allows the A^μ field to spontaneously change into the ρ' field.

There is still another trick up our sleeves. We can now pick a particular gauge. Although the form of the Lagrangian will change when we do this, its physical meaning will stay the same (that is what we mean by gauge-invariant). We change ϕ with a phase as follows: $\phi \rightarrow e^{i\theta}\phi$, where $\tan\theta = -\phi_2/\phi_1$. This particular choice makes the ρ' field disappear, but we are constrained by local gauge invariance and we are modifying the A^μ fields to compensate for this. What we get is

$$\mathcal{L} = \overbrace{\frac{1}{2}(\partial_\mu\rho)(\partial^\mu\rho) - \mu^2\rho^2}^{\text{massive scalar}} - \overbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}q^2\frac{\mu^2}{\lambda^2}A_\mu A^\mu}_{\text{massive vector}} \quad (12.19)$$

+ interaction terms

¹¹It hasn't quite disappeared—when the vector field became massive, it changed from having two polarization states to having three.

We have created a mass for the spin-1 (vector) field and the Goldstone boson has disappeared.¹¹

This is the Higgs mechanism.

The real Higgs theory is an extension of the above outline. A somewhat more complicated ‘doublet’ of Higgs scalars is used to begin with—this allows local gauge invariance to be achieved in the manner shown above with massless versions of all four fields of interest (W^\pm , W^0 , and B^0). We choose a place in the Mexican hat potential to expand around as before, and choose the gauge to make the problematic ‘ $A_\mu \partial^\mu \rho'$ ’ terms go away.

When we do this, everything works out to agree with experiment, as described in the simplified description in Section 12.3; the W^\pm and Z^0 each acquires a mass and the γ remains massless. Also, $m_W/m_Z = \cos \theta_W$ and the scalar Higgs remains massive. However, the theory does not predict the mass.

12.6 Higgs discovery

In this section, we review the evidence for the discovery of the Higgs boson. The experimental results shown come from the ‘discovery’ papers.¹² As mentioned above, the SM does not predict the mass of the Higgs boson. However, once the mass is known, all the properties are fully specified. The predicted cross section as a function of Higgs mass m_H is shown in Fig. 12.6 [114]. From the direct LEP Higgs search and the indirect constraints from the precision electroweak data, the 95% confidence level for the expected mass is $114 \text{ GeV} < m_H < 149 \text{ GeV}$ [115].

¹²For the latest results, see the ATLAS and CMS links on this book's website.

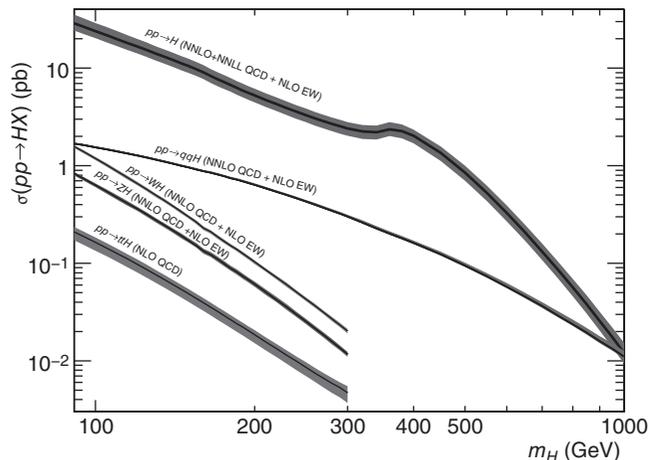


Fig. 12.6 Theoretical cross section for different Higgs production mechanisms [114] as a function of m_H for pp interactions at $\sqrt{s} = 7 \text{ TeV}$.

In this mass range, the dominant production mechanism is via a top-quark loop, because of the very large top-quark mass (see Fig. 12.7) and the large gluon parton distribution function in this x range.

The branching ratios of the Higgs boson as a function of m_H are shown in Fig. 12.8 [114]. We can see that the Higgs boson tends to decay to the heaviest particles that are kinematically allowed, since the mass of a particle depends on its coupling to the Higgs boson. In the expected range for m_H , the largest branching ratio is for decays into $b\bar{b}$ quarks; however, this channel is almost impossible to study in pp interactions in this production mode¹³ because of the large irreducible background from QCD production of $b\bar{b}$ quarks.

At lowest order in perturbation theory, the decay $H \rightarrow \gamma\gamma$ would not occur, since the photon mass $m_\gamma = 0$. However, the decay can occur through virtual t and W loops as shown in Fig. 12.9. There is significant negative interference between the two diagrams.

The most important decay channels for the Higgs boson discovery are as follows:

- $H \rightarrow \gamma\gamma$: At first sight, this might appear very surprising because the branching ratio is only about 0.2%. The invariant mass of the $\gamma\gamma$ system can be determined rather precisely, so the search in this channel then involves looking for a small ‘bump’ on top of a smoothly falling background. This has the additional benefit that the background can be determined by an empirical fit to the $\gamma\gamma$ invariant mass distribution.
- $H \rightarrow ZZ^*$: For the actual mass of the Higgs boson, this channel is below threshold for production of two real Z s; therefore, one of the Z s will be off mass shell (denoted Z^*), which reduces the branching ratio to very small values. In addition, only the decays of the Z s into electrons and muons are used. The advantages of

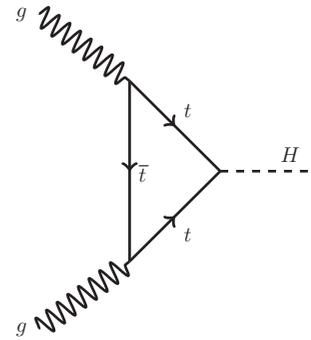


Fig. 12.7 Feynman diagram for Higgs production via a top-quark loop.

¹³An alternative production mode is VH , where V is either a W^\pm or a Z . The cross sections are smaller but this mode has much smaller backgrounds and is expected to be observed at the LHC.

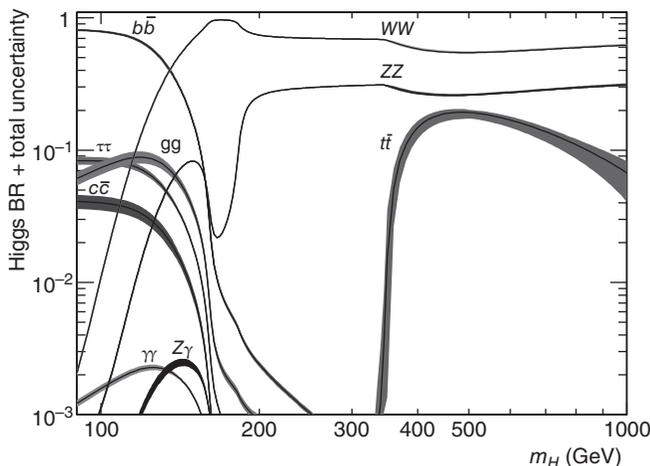


Fig. 12.8 Higgs branching ratios [114] as functions of m_H .

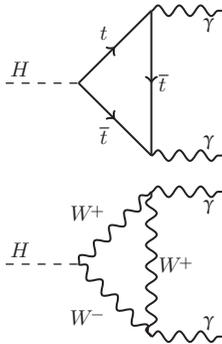


Fig. 12.9 Feynman diagrams for the decay $H \rightarrow \gamma\gamma$ via a top-quark loop (a) and a W loop (b).

¹⁴The ‘reducible’ background is one that could be removed if the detector were perfect. The ‘irreducible’ background has the same final state as the signal, and hence it is impossible to remove it on an event-by-event basis. However, in general, we can use some distributions (e.g. the invariant mass) to make a statistical separation between signal and background.

¹⁵The azimuthal angle can be determined from a line joining the shower centre to the centre of the detector. In the longitudinal direction (z) the event vertex distribution has a significant spread in z . The longitudinal angle can be determined from the shower centres in the different longitudinal layers. Alternatively, an algorithm can be used to determine which is the correct vertex (at typical LHC luminosities, there will be about 20 vertices) and the γ longitudinal angle can then be determined from a line joining the event vertex to the shower centre.

Fig. 12.10 Response of the ATLAS EM calorimeter [26] for example events: (a) $\pi^0 \rightarrow \gamma\gamma$; (b) prompt γ . The prompt γ makes a single shower, whereas the two γ s from the π^0 decay show evidence for two distinct clusters.

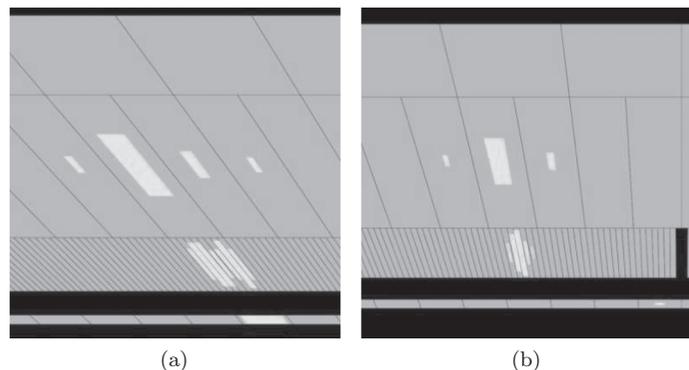
this channel are that it is very clean and also allows for a precise reconstruction of the invariant mass of the ZZ^* .

- $H \rightarrow WW^*$: This channel is also suppressed for the actual mass of the Higgs boson because one of the W s is off mass shell. The decay modes of the W that are used are $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$. The signal is larger than for ZZ^* , but as there are two neutrinos in the final state it is impossible to reconstruct an invariant mass. The transverse mass (see Section 12.6.3) is used, because the spectrum has an endpoint at the value of the invariant mass. This means that the signal peak is very broad and a very careful prediction of all the SM backgrounds is required.

12.6.1 $\gamma\gamma$ channel

We can have γ s produced in the primary interaction (called ‘prompt’) or as a result of the decays of mesons (mainly π^0). There is a potentially large reducible background¹⁴ from events with one prompt γ with the other jet faking a prompt γ or from two-jet events with both jets faking a γ . These reducible backgrounds can be suppressed with a very high-granularity electromagnetic (EM) calorimeter. $\pi^0 \rightarrow \gamma\gamma$ will tend to produce two distinct showers in the EM calorimeter as opposed to the single shower from a genuine prompt γ . In addition, the prompt γ will tend to be isolated whereas γ from π^0 decays will in general be part of a hadronic jet and will therefore not be isolated. The different responses in the ATLAS calorimeter for photons from a π^0 and a prompt γ are illustrated in Fig. 12.10. This allows the reducible background to be decreased to a level well below that of the irreducible background. The irreducible background is from the QCD production of prompt $\gamma\gamma$ (see Fig. 12.11).

The very good energy resolution of the EM calorimeter is critical for a precise reconstruction of the $\gamma\gamma$ mass $m_{\gamma\gamma}$. However, the γ directions must also be well measured for a precise reconstruction of $m_{\gamma\gamma}$ and this requires good granularity in the EM calorimeter.¹⁵ The $m_{\gamma\gamma}$ spectrum is fit to a combination of an empirical background function and a signal



distribution. The signal shape is taken from the SM prediction after simulating all detector effects, for given assumed values of m_H . Since the expected statistical significance of various classes of events (e.g. whether the γ converted into an e^+e^- pair in the tracking detector) is different, it is advantageous to plot the $m_{\gamma\gamma}$ spectrum weighted by the expected ratio of signal to background. An example of such a plot from the Compact Muon Solenoid (CMS) experiment is shown in Fig. 12.12 [64]. The plot appears to show a Higgs-like signal on top of a smooth background, but the discussion of the statistical significance will be deferred until Section 12.6.4.

12.6.2 ZZ^* channel

In order to have a clean signal and to be able to make a precise reconstruction of the mass of the Higgs boson, we use this mode with electron and muon final states: $Z/Z^* \rightarrow e^+e^-$ or $Z/Z^* \rightarrow \mu^+\mu^-$. In these decay modes, we have the best signal-to-background (S/B) ratio, but the rate is suppressed by the small branching ratios for the Z/Z^* decay modes used. The main reducible background for this channel come from events with two ‘prompt’ leptons and two from the semileptonic decays of b quarks (e.g. $Zb\bar{b}$ and $t\bar{t} \rightarrow WbW\bar{b}$). The relatively long flight path of B hadrons can be used to veto leptons from the decays of b quarks. The irreducible background is from ZZ^* production without a Higgs boson as an intermediate state, which is indistinguishable from the signal, apart from the fact that the mass spectrum of the four leptons (m_{4l}) for the signal will show a peak at m_H , whereas the background will be a smooth distribution. Clearly, the excellent energy and momentum

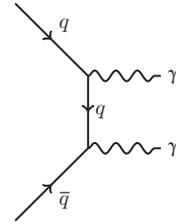


Fig. 12.11 Lowest-order Feynman diagram for prompt $\gamma\gamma$ production.

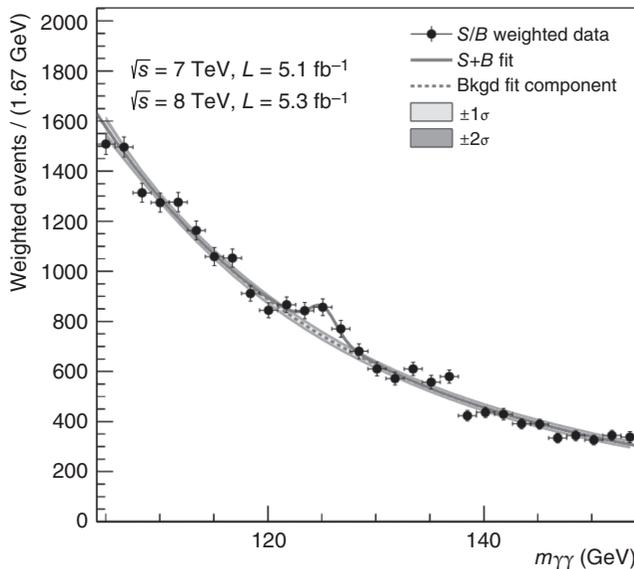


Fig. 12.12 Distribution of $m_{\gamma\gamma}$ weighted by S/B , with S and B being the expected signal and background, respectively, from the CMS experiment [64].

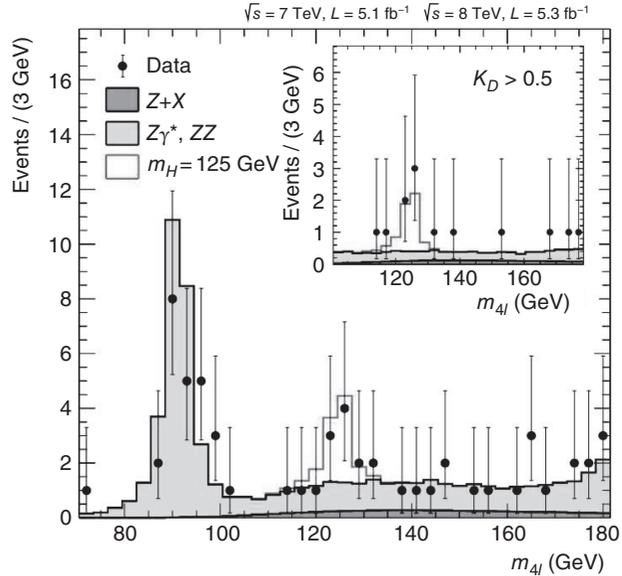


Fig. 12.13 Distribution of m_{4l} from the CMS experiment [64]. The expected background and signal for the case $m_H = 125$ GeV are also shown. The insert shows the mass distribution for a subsample of the events that pass a kinematic selection designed to optimize the ratio of signal to background.

resolutions for electrons and muons, respectively, help improve the S/B ratio. The measured distribution of m_{4l} in the CMS experiment [26] is compared with the combination of the expected background and signal for a SM Higgs (with $m_H = 125$ GeV) in Fig. 12.13.

12.6.3 WW^* channel

This channel benefits from the larger Higgs branching ratio compared with that for the ZZ^* channel. The channel that is least contaminated by background and therefore has the best significance is the one with electrons and muons: $W/W^* \rightarrow e\nu_e$ and $W/W^* \rightarrow \mu\nu_\mu$. This decay mode has larger branching ratios than the decay modes we used for ZZ^* (see Section 12.6.2), so the total number of events expected is significantly greater. The disadvantage of this channel is that there are two neutrinos in the final state, so it is not possible to determine the invariant mass (M_{WW^*}). If there were only one neutrino, the transverse mass would provide a sharp endpoint to the spectrum as in the case of single- W production. It still turns out to be useful to define the transverse mass in a similar way as for single- W production:

$$m_T^2 = (E_T^{ll} + E_T^{\text{miss}})^2 - (\mathbf{p}_T^{ll} + \mathbf{E}_T^{\text{miss}})^2 \quad (12.20)$$

where \mathbf{p}_T^{ll} is the total momentum in the transverse plane of the two charged leptons, E_T^{miss} and $\mathbf{E}_T^{\text{miss}}$ are respectively the magnitude of the missing transverse momentum and the missing transverse momentum vector, and

$$(E_T^{ll})^2 = (\mathbf{p}_T^{ll})^2 + m_{ll}^2 \quad (12.21)$$

where m_{ll} is the invariant mass of the two charged leptons. It turns out that the distribution of m_T has an endpoint at the value of the invariant mass of the Higgs boson. However, the effective mass resolution is only $\sim 20\%$. There are significant reducible backgrounds from $W + \text{jet}$ events with the jet faking a lepton and from Drell–Yan processes (see Chapter 9) with additional jets (e.g. $e^+e^- \text{ jet jet}$ or $\mu^+\mu^- \text{ jet jet}$) with fake missing transverse energy from mismeasurements. Therefore, in ATLAS, only events with one W decaying to $e\nu_e$ and one decaying to $\mu\nu_\mu$ were considered. There remains an irreducible background from WW production. In order to reduce this background on a statistical basis, various cuts are used. For example, for WW arising from the decay of a spin-0 Higgs, the spins of the two W s must point in opposite directions and the two charged leptons will tend to have momenta in similar directions (see Exercise 12.11). The irreducible background from WW production (see Fig. 12.14 for an example Feynman diagram for WW production) must be determined very precisely in order to be able to detect a significant excess from Higgs events. This is done by selecting events with different kinematic cuts to obtain background-dominated samples. This is used to fix the normalization of the background. The ratio of the number of expected background events in the signal region to the number of events in the background (‘control’) region is taken from Monte Carlo simulation. The transverse mass distribution measured by ATLAS is shown in Fig. 12.15 [26, 30] and compared with the expectations from background and a SM Higgs signal for $m_H = 126 \text{ GeV}$.

12.6.4 Statistical significance

The relatively small signals in all channels and the non-negligible backgrounds mean that a careful assessment of the statistical significance is required before any claims about a discovery can be made. The technique

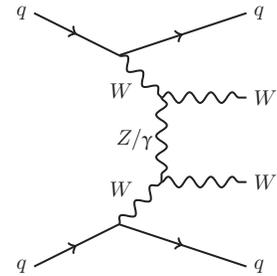


Fig. 12.14 An example Feynman diagram at the quark level for WW scattering.

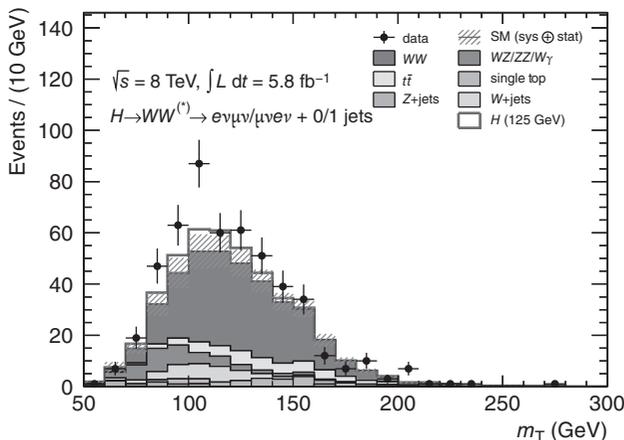


Fig. 12.15 Distribution of the transverse mass m_T from the ATLAS experiment [26] in a sample of candidate WW events. The expected background and signal for the case $m_H = 126 \text{ GeV}$ are also shown.

¹⁶By ‘shape’ we mean the shape of the distributions for the relevant variables.

¹⁷In practice, we use a ‘binned’ likelihood, in which the events are binned in the relevant variable(s) and the product is over the bins. In addition, for ease of computation, we work with a ‘log-likelihood’ $\ln L(\mu) = \sum_i \ln p(i | \mu, \theta)$.

¹⁸In order to simplify the explanation, we now ignore the nuisance parameters.

used is based on maximum-likelihood fits. Probability density functions (PDFs) are computed assuming there is a signal with the same shape¹⁶ as the SM Higgs but scaled by a normalization constant μ (i.e. $\mu = 1$ corresponds to the SM Higgs expectation and $\mu = 0$ corresponds to a pure background distribution). The PDF also depends on many ‘nuisance’ parameters (such as uncertainties in the detector response to different particles). The likelihood of a sample¹⁷ is defined as

$$L(\mu) = \prod_i p(i | \mu, \theta) \quad (12.22)$$

where the product runs over all events in the sample and $p(i | \mu, \theta)$ is the probability of observing event i , assuming a particular value of the parameter μ and for some set of nuisance parameters θ . The test statistic is defined as¹⁸

$$\tilde{q}_\mu = -2 \ln \left[\frac{L(\text{data} | \mu)}{L(\text{data} | \hat{\mu})} \right] \quad (12.23)$$

where $\hat{\mu}$ is the value obtained by maximizing the likelihood by varying the value of μ in the denominator in eqn 12.23 and the value of μ in the numerator is fixed to a particular value depending on what statistical test is being performed. We can then define two probabilities:

- For signal plus background, $p_\mu = P(\tilde{q}_\mu > \tilde{q}_\mu^{\text{obs}} | \mu)$.
- For the background only, $1 - p_b = P(\tilde{q}_\mu > \tilde{q}_\mu^{\text{obs}} | \mu = 0)$.

Finally, we define the ratio

$$\text{CL}_s(\mu) = \frac{p_\mu}{1 - p_b} \quad (12.24)$$

The 95% confidence level (CL) limit on μ is found by adjusting μ until $\text{CL}_s(\mu) = 0.05$. This procedure is carried out for a range of values of m_H and we can state the SM Higgs boson is excluded at 95% confidence level for a particular value of m_H if $\mu(95\% \text{CL}) < 1$.¹⁹ The resulting exclusion plot from ATLAS is shown in Fig. 12.16(a) [26]. The SM Higgs is excluded over the ranges 111–121 GeV and 131–559 GeV. The reason why there is a gap in the exclusion range between 121 and 131 GeV is that there is evidence for an excess over the SM background. The probability of this excess can be quantified by the probability p_0 that a background-only hypothesis could generate a larger value of the test statistic q_μ for a particular value of m_H . The resulting p_0 versus m_H plot for ATLAS is shown in Fig. 12.16(b). The minimum p_0 value corresponds to a 6σ fluctuation for a Gaussian distribution. However, it is important to allow for the ‘look-elsewhere effect’; the p_0 value corresponds to the probability of observing a larger fluctuation at a particular value of m_H , but we could have seen an excess over a range of values for m_H . Therefore, the probability of observing an equal excess over the full mass range is larger by a factor $N \sim \Delta m_H / \sigma(m_H)$, where Δm_H is the range in m_H and $\sigma(m_H)$ is the resolution in the event-by-event measurement.

¹⁹In the frequentist interpretation of statistics, the statement that the SM Higgs is excluded at a given value of m_H at 95% confidence level means that if there were a SM Higgs at this mass and the experiment were repeated many times, then in 95% of the cases a larger value of the test statistic would be obtained.

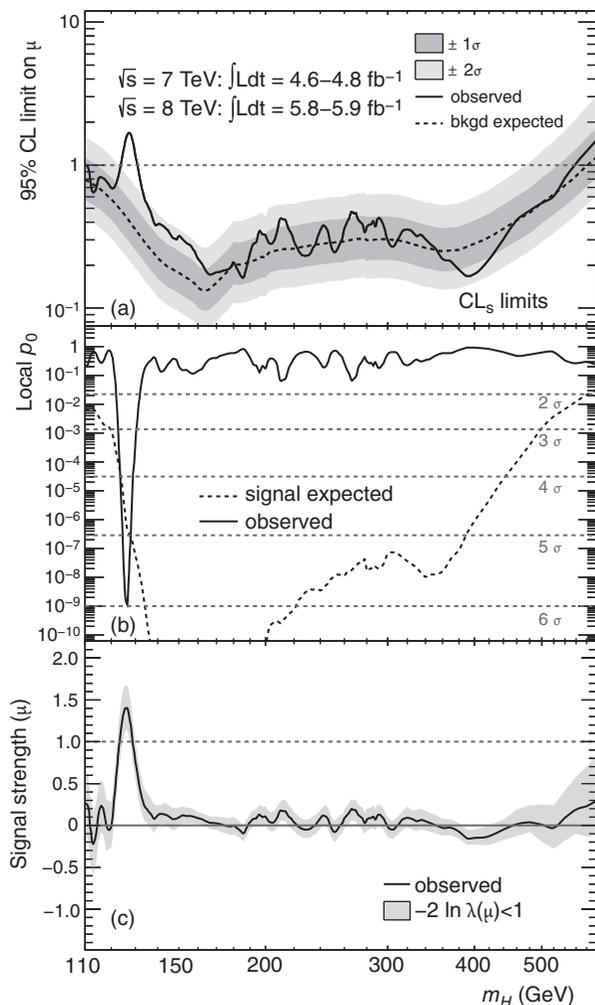


Fig. 12.16 (a) 95% confidence limit on the SM Higgs. (b) Local p_0 values as a function of m_H . (c) Variation of fitted signal strength μ with m_H from the ATLAS experiment [26].

Allowing for this effect for a search over the allowed range from previous experiments at LEP and the Tevatron (110–150 GeV), the statistical significance is reduced to 5.3σ .²⁰ A very similar result was obtained from the CMS experiment [64].²¹ Finally, one can ask if the observed excess is consistent with the SM Higgs. This is addressed in Fig. 12.16(c), which shows the 95% confidence level on the signal strength parameter μ as a function of m_H . This shows that the result is consistent with the SM Higgs expectations for a Higgs with $m_H = 126 \text{ GeV}$.

12.7 Coupling to fermions

The SM Higgs mechanism generates masses for fermions as well as for bosons. Therefore, the Higgs boson should have decay modes to fermions

²⁰The look-elsewhere effect only accounts for the range of mass for this one study. Since experiments make many independent measurements, we want to minimize the probability of falsely claiming a discovery, so a high threshold in significance is used before claiming a discovery. Conventionally, this is taken to be 5σ .

²¹Consistent results were obtained by the Tevatron experiment at the 3σ level, mainly in the $b\bar{b}$ channel.

²²This decay mode should be observable at the LHC in the VH production mode, where $V = W^\pm$ or Z . However, as this production mode has a smaller cross section, it will need larger data samples than are currently available (as of 2014).

²³The backgrounds are so large that simply making a sequence of cuts to optimize signal and reduce background would not be sufficient to reveal a signal, and so a more sophisticated multivariate analysis is required (see Chapter 8).

that grow with the mass of the fermion. As the Higgs boson is too light to decay into $t\bar{t}$, the most massive fermion available for this process is the b quark and hence there is a large branching ratio for $H \rightarrow b\bar{b}$. However, this decay mode is extremely difficult to study because of the very large QCD production of $b\bar{b}$.²² The next heaviest fermion in the SM is the τ lepton and there is a significant branching ratio for $H \rightarrow \tau\bar{\tau}$ (see Fig. 12.8). The identification of τ leptons is much harder than the identification of e or μ . There are many backgrounds to consider. The $H \rightarrow \tau\tau$ channel has an irreducible background from $Z \rightarrow \tau\tau$. Separation between the signal and this background can be achieved by reconstructing the mass of the Higgs boson, but the mass resolution is severely degraded by the presence of multiple neutrinos in the final state (see Section 12.6.3). The leptonic decay modes of the τ have smaller branching ratios than the hadronic modes. However, the hadronic decays of the τ have very large backgrounds from QCD jets that can be misinterpreted as τ s. Some separation between hadronically decaying τ s and QCD jets is obtained by counting the number of charged tracks in a narrow cone around the τ candidate. The number of tracks from τ decays is usually one or three, whereas QCD jets have a higher average charge multiplicity. Additional separation between τ hadronic decays and QCD jets is provided by measurements of the shower profile in the calorimeter, since the τ s will tend to produce relatively narrow jets.²³ Although the invariant mass cannot be reconstructed, a best estimate of the mass can be made (see Exercise 12.10 for a simpler method to estimate the Higgs mass in τ decays). The mass distribution for the data and all the SM backgrounds are shown in Fig. 12.17 and a peak above

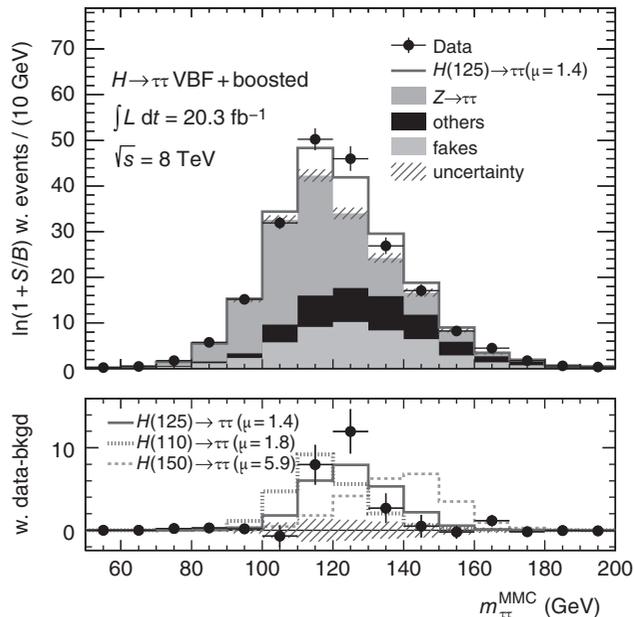


Fig. 12.17 Distribution of the best estimate for the Higgs boson mass for data and SM backgrounds for candidate $\tau\tau$ events. All events have been weighted by a factor $\ln(1 + S/B)$, where S/B is the expected signal-to-background ratio [28].

the SM background can be seen. The peak is consistent with the SM expectation and provides evidence at the 4.1σ level for this decay mode of the Higgs boson [28]. The CMS experiment also found evidence for this decay mode [66].

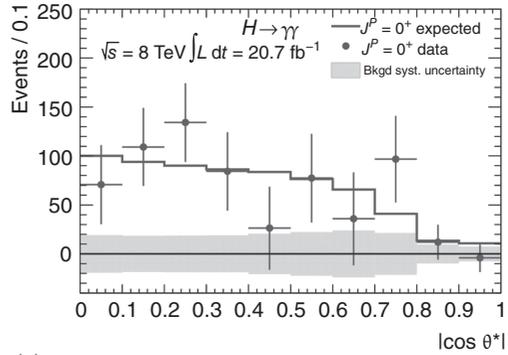
12.8 Determination of the spin and parity of the new boson

One fundamental prediction of the SM is that the spin/parity of the Higgs boson should be $J^P = 0^+$. Since the new boson decays to two bosons, it must have integer spin. The observation of the decay mode $H \rightarrow \gamma\gamma$ excludes the spin-1 hypothesis [95]. The SM predictions can be compared with models in which the boson has alternative spin/parity assignments. For the $\gamma\gamma$ decay mode, we can compare the SM with predictions for graviton-inspired models²⁴ with $J^P = 2^+$. We can define the angle of the photons relative to the direction of the Higgs boson (θ^*). The distribution of $\cos\theta^*$ should be isotropic for the case of a spin-0 Higgs. However, in the case of a boson with $J^P = 2^+$, the distribution can be forward-peaked. The angular distribution of the observed events is biased by the acceptance of the detector and the selection required to isolate the signal above the background. However, there are still significant differences between the predictions of the two models, and the data can be used to discriminate between them. The measured distribution of $\cos\theta^*$ is compared with the ATLAS data [29] in Fig. 12.18. Information on the spin/parity can also be obtained from the ZZ^* decay mode, although the errors are still quite large because of the limited number of events in the current data sample. The WW decay modes for the case of a spin-0 Higgs should result in correlations between the azimuthal angle of the charged leptons from the resulting decays of the W bosons (see Exercise 12.11). Combining the analyses from three decay modes, all the results are consistent with the SM quantum numbers of $J^P = 0^+$ and alternative models can be excluded at a range of confidence levels from 97.7% to 99.9% [28]. Results of a similar analysis from the CMS experiment [65] in the $H \rightarrow ZZ$ decay mode also favour the SM and disfavour alternative models.

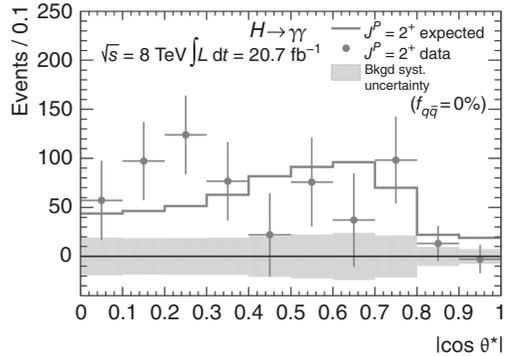
A clear summary of the Higgs boson coupling measurements from the CMS experiment [67] is shown in Fig. 12.19. This shows the measured strength of the Higgs boson coupling to fermions (λ) or bosons (g) as a function of the mass of the particle. For the top quark, the decay $H \rightarrow t\bar{t}$ is kinematically forbidden because of the large mass of the top quark. However, we can still determine the coupling of the top quark to the Higgs boson, because the dominant production mechanism is via a top-quark loop (see Fig. 12.7). In the SM, we expect the coupling strengths to increase with the mass of the particle. The measured coupling strengths show the expected increase with particle mass as expected in the SM. However, the errors in some of the measurements are large and more data will make this test more powerful.²⁵

²⁴The graviton is a hypothetical particle that is believed to be the quantum of the gravitational field. The graviton must have spin 2 to correspond to the classical theory of general relativity, in which the gravitational interaction arises from the stress-energy tensor, which is a second-rank tensor.

²⁵Eventually, there should be sufficient data at the LHC to see the $H \rightarrow \mu^+\mu^-$ decay mode and thus provide an additional data point at much lower mass.



(a)



(b)

Fig. 12.18 Angular distribution of $\cos \theta^*$ measured by ATLAS [29] compared with two hypotheses for the spin/parity: (a) $J^P = 0^+$; (b) $J^P = 2^+$.

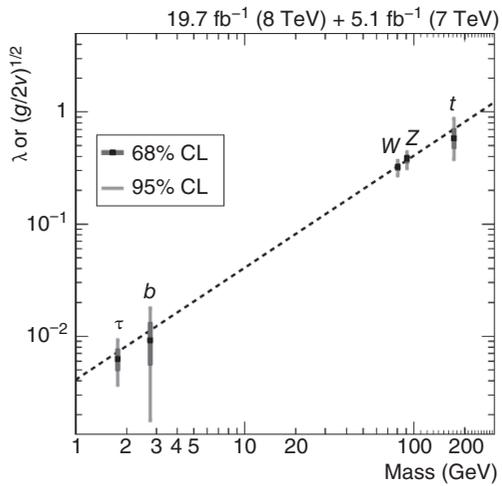


Fig. 12.19 Measured values of the Higgs boson coupling to fermions (λ) or bosons (g) as a function of particle mass from the CMS experiment [67].

12.9 Outlook

The results already obtained provide convincing evidence for the existence of a new boson (it decays to $\gamma\gamma$, so it must have even spin).²⁶ The study of the spin/parity gives results consistent with the SM expectation of 0^+ and other possibilities are excluded. The properties appear to be consistent with the SM Higgs, but this is just the beginning of a new chapter in physics, which will involve many more detailed measurements:

- (1) a precise measurement of the rate to check compatibility with the SM;
- (2) measurements of ratios of branching ratios for as many decay modes as possible;
- (3) measurements of the ratio of production cross sections for gg processes and vector-boson fusion (e.g. $qq \rightarrow qqWW \rightarrow qqH$);
- (4) measurements of high-energy longitudinally polarized WW scattering; this process is sensitive to additional scalars in the theory or to the possible composite nature of the Higgs boson;
- (5) Higgs self-interactions, i.e. production of two Higgs bosons in one event.

The first three items will be studied with new data in the next few years at the LHC. The last two items are more challenging and will require a further upgrade to the LHC luminosity.²⁷ In the SM, there is only expected to be one Higgs boson, but in many Beyond the SM (BSM) theories, there can be more. For example, in the minimal supersymmetric extension of the SM (MSSM), there should be five Higgs bosons (see Chapter 13). Therefore, searching for additional Higgs bosons will clearly be a vital part of the future LHC physics programme.

²⁶The $\gamma\gamma$ decay mode excludes spin 1.

²⁷Such an upgrade to the LHC is planned for 2024. This will also require extensive upgrades to the ATLAS and CMS detectors to cope with the higher luminosity.

Chapter summary

- The Higgs mechanism allows for the generation of mass by spontaneous symmetry breaking, without violating the $SU(2) \times U(1)$ gauge symmetry.
- There is compelling evidence for the existence of a new boson at a mass of 125 GeV, with properties consistent with those expected from the SM.
- The measurements of the spin/parity favour the 0^+ assignment expected in the SM.
- Many more measurements will be required to confirm that the new boson is indeed the SM Higgs boson. Searches will be made for additional Higgs bosons as expected in theories like supersymmetry.

Further reading

- Griffiths, D. (2008). *Introduction to Elementary Particles* (2nd revised edn). Wiley-VCH. This gives a more advanced treatment of the Higgs mechanism.
- Burcham, W. E. and Jobes, M. (1994). *Nuclear and Particle Physics*. Pearson. This gives a fuller but accessible account of the Higgs mechanism in the SM.
- Aitchison, I. J. R. and Hey, A. J. G. (2013). *Gauge Theories in Particle Physics*, Volume 2 (4th edn). CRC Press. This is a thorough and more advanced graduate-level treatment of the Higgs mechanism.

Exercises

- (12.1) Apply the gauge transformations for the \mathbf{A} , ϕ , and ψ fields in eqn 12.3 and hence verify that the Dirac equation is gauge-invariant.
- (12.2) Evaluate the potential and kinetic energy terms for a spherical pendulum and hence verify the form of the Lagrangian in eqn 12.7. Using the Euler–Lagrange equations and this Lagrangian, determine the equations of motion and check that they agree with eqn 12.8.
- (12.3) A Taylor series. Show that the following three expressions are identical:

$$\begin{aligned}
 f(x) &= x^4 - 2x^2 \\
 f(x) &= -1 + 0(x-1) + 4(x-1)^2 + 4(x-1)^3 \\
 &\quad + (x-1)^4 \\
 f(x) &= -1 + 0(x+1) + 4(x+1)^2 - 4(x+1)^3 \\
 &\quad + (x+1)^4
 \end{aligned}$$

The second and third expressions are in the form of Taylor series of the first about the points $x = 1$ and $x = -1$, respectively, which are the two minima of the function. The series terminate (why is this?) and so we end up with terms only up to the fourth order.

- (12.4) Verify that applying the Euler–Lagrange equation to the Lagrangian for a spin-0 particle (eqn 12.11) gives the Klein–Gordon equation.

Note: If your answer has an unwanted factor 2, check by first expanding the $(\partial_\mu\phi)(\partial_\nu\phi)$ term into the four components.

- (12.5) Consider the Lagrangian given by

$$\mathcal{L} = -(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu)$$

Apply the Euler–Lagrange equations to show that the equations of motion are $\partial_\mu F^{\mu\nu} = 0$, where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Comment on the relation of this result to classical electromagnetism.

- (12.6) Consider the potential energy function given by eqn 12.4. Show that $\phi \pm \mu/\lambda$ correspond to the minima of the potential and $\phi = 0$ is the maximum. Let $\rho = \phi \pm \mu/\lambda$ to express the potential as a function of ρ . Show that this function now has a positive mass term and determine the value of the mass. We now have a ρ^3 term. Draw the corresponding Feynman diagram. We obtain a similar term in the full Higgs theory; explain the significance in terms of di-Higgs production.
- (12.7) Figure 12.6 shows an enhancement of the Higgs boson production cross section as a function of mass at a mass around 350 GeV. Give a simple explanation for this.
- (12.8) Figure 12.8 shows the branching ratio as a function of Higgs mass for Higgs boson decays to ZZ . Give a simple explanation for the behaviour as a function of Higgs mass.
- (12.9) The expected background in the four-lepton invariant mass (see Fig. 12.13) shows peaks around 90 and 200 GeV. Give a qualitative explanation for these peaks. Using the data in Fig. 12.13 and the theoretical value for the branching ratios (see Fig. 12.8), make an order-of-magnitude

estimate of the Higgs boson production cross section. Explain whether your estimate would be an underestimate or an overestimate.

- (12.10) This question looks at the reconstruction of the Higgs boson mass in $H \rightarrow \tau\tau$ decays in the collinear approximation. The momenta of the neutrino(s) from each τ decay are assumed to be parallel to each τ . Show that if the angles in the transverse plane for each τ are measured, and the transverse momentum of the recoiling hadronic system (\mathbf{R}) is measured, we can determine the transverse momenta of the two τ s. What other parameters must be measured in order to determine the invariant mass of the Higgs boson? Why does this method break down if

the Higgs boson is produced with no transverse momentum?

Hint: It is sufficient to consider the special case in which the two τ s decay along the y axis.

- (12.11) Consider the decay of a spin-0 Higgs to W^+W^- with subsequent leptonic decays of the W s. Let $\Delta\phi$ be the azimuthal angle between the two final-state leptons. Draw diagrams to illustrate the possible polarization states of the W s. What would be the most probable decay angle of the leptons relative to the spin of the W s? Hence explain why the distribution of $\Delta\phi$ should peak at small values.

Why do the experiments use the relative angle between the leptons in the transverse plane ($\Delta\phi$) rather than the space angle?

13

LHC and BSM

13.1 LHC and Standard Model physics	356
13.2 LHC triggers	358
13.3 SM measurements at the LHC	359
13.4 Beyond the Standard Model physics	367
13.5 Experimental searches for BSM physics at the LHC	373
13.6 Linear collider	379
13.7 Dark matter	381
13.8 Dark energy	385
Chapter summary	387
Further reading	388
Exercises	388

Now that there is growing evidence for the discovery of the Higgs boson, there are no missing particles in the Standard Model (SM). There are still many details to be checked to be sure that the new boson discovered at the LHC (see Chapter 12) is compatible with the SM Higgs boson. Even if this turns out to be the case, however, serious problems remain with the SM. A brief review of these issues, particularly the hierarchy problem, will be given in this chapter. We will justify the claim that we should expect to see new Beyond the Standard Model (BSM) physics in the TeV energy range. This then provides one of the two principal motivations for the LHC (the first being understanding the origin of mass), and we will look at how LHC experiments are searching for new physics. The study of the Higgs boson could be extended at a future linear e^+e^- collider (a ‘Higgs factory’). Looking further into the future, we will see how a high-energy linear e^+e^- collider could probe deeper into any new physics that might be discovered at LHC, if it were at an accessible energy.

13.1 LHC and Standard Model physics

The main motivation for the LHC is to understand the origin of mass and to search for new physics at the TeV scale. The origin of mass has been discussed in Chapter 12, so we will now consider BSM physics. However, any new physics channel will have some SM physics backgrounds, so it is essential to check that the SM works in the new region of phase space opened up by the LHC. This is non-trivial because, even with the LHC operating at 8 TeV CMS energy (rather than the design value of 14 TeV), high- Q^2 processes such as W/Z production¹ sample very much lower values of the parton x distribution than previous experiments (see Chapter 9). This is illustrated in Fig. 13.1, from which one can see the large increase in the range of x that can be studied at high Q^2 at the LHC, compared with previous accelerators [115].

The predicted cross sections (see Chapter 9 for an explanation of how these calculations are performed) are shown as a function of CMS energy in Fig. 13.2. There is an enormous spread in the magnitudes of the cross sections for SM processes and most of them are much larger than those expected for Higgs production or for new physics processes.

This raises the critical question of how one can trigger and identify the interesting events in the presence of these enormous backgrounds.

¹ Q^2 is the scale of the process; e.g. for W production, $Q^2 = M_W^2$.

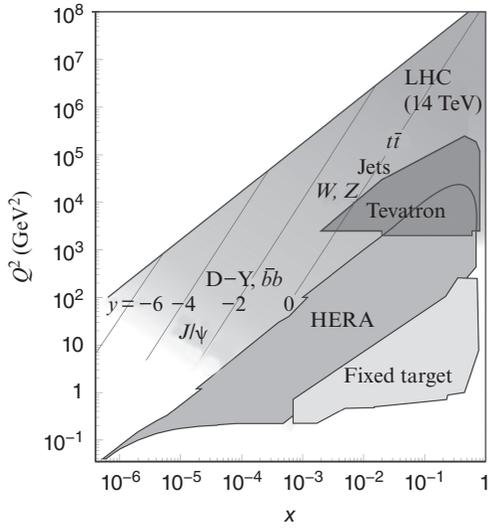


Fig. 13.1 Kinematic plane in x and Q^2 for the LHC at $\sqrt{s} = 14$ TeV, compared with HERA, Tevatron, and fixed-target experiments [115].

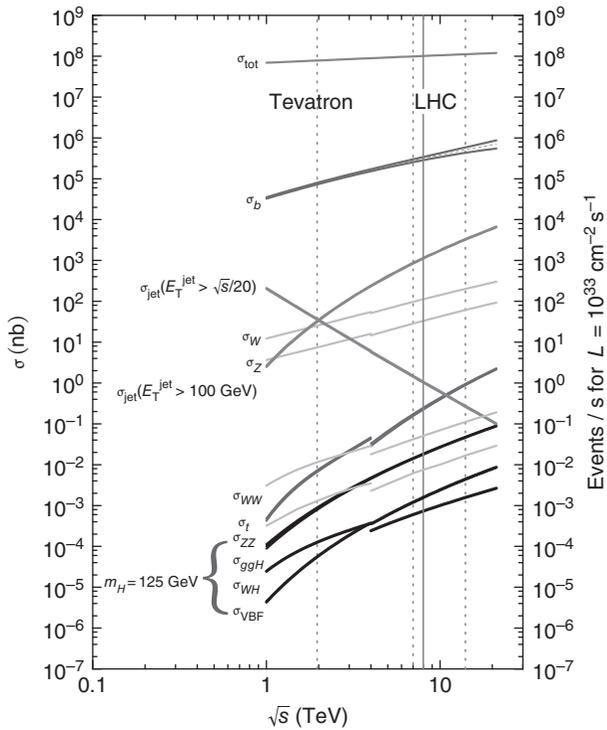


Fig. 13.2 Theoretical cross sections for different processes as functions of CMS energy for $p\bar{p}$ (Tevatron) and pp (LHC). Figure from private communication from Professor James Stirling.

13.2 LHC triggers

The scale of the problem is set by the magnitude of the total cross section. At the LHC design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the total interaction rate is $\sim 1 \text{ GHz}$. Each proton beam is composed of ~ 2000 bunches of protons. Each bunch contains $\sim 10^{11}$ protons. During nominal LHC operation, the bunches collide every 25 ns and there are an average of 24 interactions per bunch crossing (b.c.).² The general principles of pipelined triggers were briefly reviewed in Chapter 4. Most of the triggers are based on identification of objects like electrons, photons, muons, or ‘jets’ at high transverse momentum p_T . An example of a first-level (L1) trigger in ATLAS is the e/γ trigger, which uses the fact that electrons give very localized energy deposition in the electromagnetic (EM) calorimeter, whereas the background processes arise from QCD jets, which consist of mixtures of γ s (mainly from π^0 decays) and hadrons. The L1 e/γ trigger selects candidates by requiring localized energy in the EM calorimeter and vetoing events with too much energy in the hadronic calorimeter behind the candidate e/γ object or in the surrounding cells of the EM calorimeter. The higher-level triggers can utilize data from the tracking detector to provide further background rejection. A genuine electron will have a track with a transverse momentum and direction compatible with the energy deposition in the EM calorimeter, unlike the background from QCD jets. The fine granularity of the EM calorimeter is also used to provide more background rejection because EM showers are narrower than hadronic showers (see Chapter 4).

The very large trigger rejection must be achieved while still maintaining high efficiency for the interesting objects like electrons and muons. This raises the crucial question as to how one can determine if the trigger system is working efficiently or not. In general, this can be done using a ‘tag-and-probe’ analysis. For example, $Z \rightarrow e^+e^-$ events can be triggered with a single electron trigger on one of the electrons in the event. One of the electrons (we do not distinguish between electrons and positrons for this analysis) has a reconstructed electron passing the electron selection and matching in the detector location with the trigger, and this electron serves as the ‘tag’ for the event. We can then look for a second electron in the event that also passes the electron selection, and this electron serves as the ‘probe’. We can examine the trigger data and determine if this electron also passed the electron trigger. Let n_{tag} be the number of events in which an event is tagged and n_{probe} the number of events in which the second object, the ‘probe’, also passes the trigger. The trigger efficiency can be determined to be (see Exercise 13.1)

$$\epsilon_e = \frac{2n_{\text{probe}}}{n_{\text{probe}} + n_{\text{tag}}} \quad (13.1)$$

The beauty of the tag-and-probe method is that it provides a purely data-driven efficiency determination and so does not rely on any assumptions about the detector performance that are needed for a Monte Carlo simulation. This method can be extended from a ‘global’ measurement

²In 2012 LHC, operation, the luminosity was slightly lower than the design value, but the b.c. spacing was 50 ns, so the average number of interactions per bunch crossing was ~ 30 .

to a differential one in which the efficiency is measured as a function of variables like the transverse momentum p_T .³

We now consider the critical issue of how it is possible to search for very rare processes, despite the large backgrounds. The processes with the largest cross sections result in particles with limited transverse momentum and can easily be rejected by any trigger requiring a high- p_T object. For example, jet events can be studied by triggering on large transverse energy in a localized region of the calorimeter. In order to trigger on rarer processes such as W and Z production, one can trigger on leptons. For the very small cross sections for processes like Higgs production, triggers on multiple objects can be used to further reduce the trigger rate. In addition to triggers on localized objects, there is also a very important trigger on ‘missing transverse momentum’ to detect weakly interacting particles like neutrinos. Let the energy measured in a calorimeter cell (assuming massless energy deposits, so $E = p$) be E_i and let the polar and azimuthal angles be θ_i and ϕ_i . The measured momentum balance in the plane perpendicular⁴ to the beam axis (z) is defined by

$$\begin{aligned} E_{T,x}^{\text{miss}} &= - \sum_i E_i \sin \theta_i \cos \phi_i \\ E_{T,y}^{\text{miss}} &= - \sum_i E_i \sin \theta_i \sin \phi_i \end{aligned} \quad (13.2)$$

and the magnitude of the missing transverse momentum is given by

$$E_T^{\text{miss}} = \sqrt{(E_{T,x}^{\text{miss}})^2 + (E_{T,y}^{\text{miss}})^2} \quad (13.3)$$

This E_T^{miss} trigger is useful for selecting events with neutrinos but is also essential for searches for BSM physics such as supersymmetry that have weakly interacting particles (see Section 13.4.1).

13.3 SM measurements at the LHC

Many measurements of different SM physics processes have been performed at the LHC. This section will give a brief summary of a few of these studies. The largest cross sections for high- p_T processes are for dijet production, since this is governed by the strong interaction. The data for the distribution in jet transverse momentum are compared with next-to-leading-order (NLO) QCD predictions [24] in Fig. 13.3.⁵ There is very good agreement between data and the QCD prediction up to $p_T \sim 1$ TeV. This very impressive agreement spans a dynamic range of 10 orders of magnitude. As the jet cross section is a steeply falling function of the jet transverse energy p_T , a relatively small error in the p_T measurement will result in a large error in the cross section. It is therefore essential to make an accurate calibration of the jet energies.

³The same tag-and-probe technique can also be used to determine efficiencies for offline electron identification. The methodology is very powerful since it can be extended to any process that provides two independent objects to select.

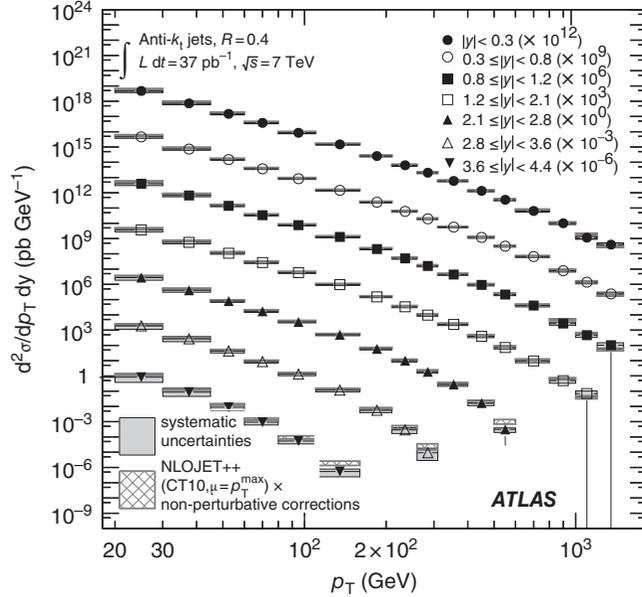
⁴In the beam direction, so much momentum is carried by particles that are ‘lost’ down the beam pipe that we cannot make a useful measurement.

⁵The QCD calculations are performed using perturbation theory, utilizing the fact that at large transverse momentum, Q^2 , the strong coupling constant $\alpha_s(Q^2) \ll 1$. The NLO calculation includes Feynman diagrams with one extra power of $\alpha_s(Q^2)$ than the leading-order diagram.

Fig. 13.3 Jet cross section versus transverse energy, measured by the ATLAS experiment at the LHC. The measurements are shown in slices of the rapidity variable

$$y = \frac{1}{2} \frac{E + p_z}{E - p_z}$$

where E is the energy and p_z is the momentum component along the beam direction for the jet. The data are compared with the predictions from NLO QCD (light-shaded bands at each data point) [24].



⁶The main source of photons is from the decays of particles like π^0 and η .

⁷‘Underlying’ event refers to the interactions of the spectator partons left behind after the hard parton-parton interaction.

⁸‘Pile-up’ refers to additional pp collisions that occur in the same bunch crossing as a triggered event.

⁹ $\Delta R = \sqrt{(\Delta\phi^2 + \Delta\eta^2)}$, where the pseudorapidity $\eta = -\ln(\tan \frac{1}{2}\theta)$ and θ and ϕ are the polar and azimuthal angles of the cell.

Jets are composed of hadrons as well as photons.⁶ The energy determination of hadrons in a calorimeter is more difficult than that of electrons or photons (see Chapter 4). In addition, there are uncertainties in the reconstruction of the energy of a hadron jet, because any jet-finding algorithm has to determine which energy depositions should be considered as part of the jet. Therefore, there will be energy depositions that fail to be counted as part of the jet and some energy depositions from the ‘underlying’ event that are wrongly attributed to the jet.⁷ At high luminosity, there is also the problem that some energy from the ‘pile-up’ events will also be wrongly associated with the jet.⁸

The simplest jet finder is based on the ‘cone’ algorithm. The highest- p_T cell in the calorimeter is used as a ‘seed’ and the nearest cell with transverse energy above some fixed threshold and within a radius ΔR less than a fixed size is found.⁹ An updated value of the jet direction is computed as the energy-weighted average of the two cells. The procedure is then iterated until no more cells are found to merge. The next jet is found starting from the seed of the remaining cell with the highest transverse energy. The procedure terminates when there are no unused cells above threshold.

While the algorithm is simple from an experimental point of view, it has major theoretical problems that make comparisons with QCD predictions difficult. Consider an event with two jets that are separated and should be reconstructed as two separate jets. If there is a ‘soft’ particle between the two partons, this can cause the two jets to merge. This makes the results very sensitive to low-energy radiation, and the algorithm is described as not being ‘infrared-safe’.

Several infrared-safe algorithms have been proposed based on sequential recombination. We define a distance d_{ij} between cells or jets i and j .

We define a similar distance d_{iB} from a cell or a jet to the beam. The clustering starts with the smallest distance d_{ij} or d_{iB} . If $d_{ij} < d_{iB}$, it combines the entities i and j , otherwise the entity is called a jet and its cells are removed from the list. The algorithm is iterated until there are no more jets to be found.

Different algorithms use different definitions of distance. The most commonly used algorithm is called the anti- k_T jet finder [58]:

$$\begin{aligned} d_{ij} &= \min \{ (p_{T,i}^{-2}, p_{T,j}^{-2}) (\Delta R_{ij}/R)^2 \} \\ d_{iB} &= p_{T,i}^{-2} \end{aligned} \quad (13.4)$$

where ΔR is the same separation in rapidity and azimuthal angle as used for the cone algorithm, and $p_{T,i}$ is the transverse momentum. R is a fixed radius in (y, ϕ) space (similar to the fixed cone radius).¹⁰

¹⁰Typical values are in the range $\Delta R = 0.4$ – 0.7 .

There are many contributions to the uncertainties in the measured energy of jets, so it is vital to use data-driven techniques to calibrate the jet energy scale. The following are some of these techniques:

- p_T balance in γ -jet events. As the γ transverse energy can be measured very accurately in the EM calorimeter, the ratio $p_T^{\text{jet}}/p_T^\gamma$ can be used to calibrate the jet energy scale. An example of a calibration plot from this analysis [37] is shown in Fig. 13.4.
- A similar technique can be used in events with Z s recoiling against a hadronic jet. The well-measured transverse momentum of the Z can be used to calibrate the jet p_T .

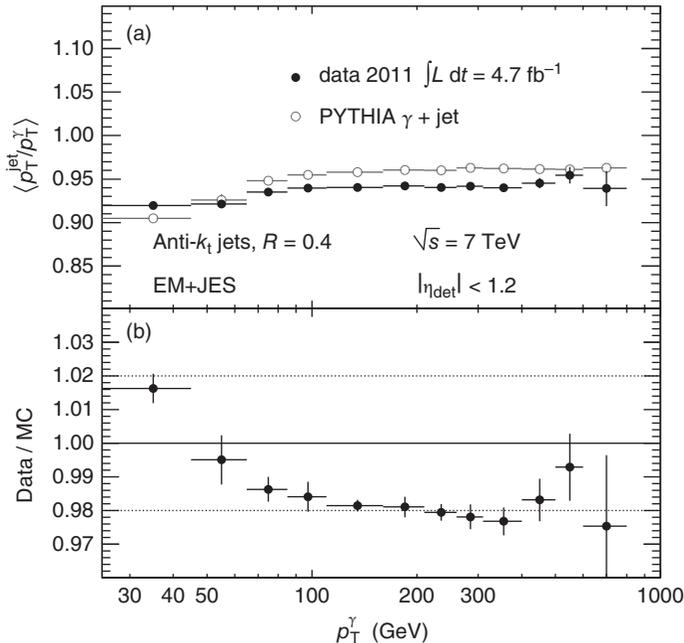


Fig. 13.4 Jet energy calibration in ATLAS using the γ -jet balance technique [37]. (a) Ratio of jet to photon transverse energies for data and Monte Carlo simulation. (b) Ratio of the values from data and Monte Carlo calculations ('PYTHIA'). Note the different scales for the two plots.

- Dijet balance: in two-jet events, the p_T of the two jets should be equal in magnitude. This cannot provide an absolute calibration, but it is a powerful tool for inter-calibration of different regions of the calorimeter.
- Multijet balance: in three-jet events, if the calibration factors for the two lowest- p_T jets are known, the p_T balance can be used to calibrate the highest- p_T jet.¹¹

¹¹This is particularly useful because there are too few events at very high p_T to use the absolute calibration techniques. As multijets are produced by purely strong interactions, the rates are higher than for γ -jet or Z -jet events.

Production of W and Z bosons with subsequent leptonic decays provides very clean channels for comparisons with QCD predictions. The first comparisons we consider are measurements of the cross sections for these processes. The methodology for the theoretical calculation in the parton model has been described in Chapter 9. For an ideal detector with 100% efficiency and no background, the measured cross section would simply be related to the number of events observed, N_{obs} , and to the integrated luminosity L by

$$N_{\text{obs}} = L\sigma \quad (13.5)$$

In a real detector, we have to subtract the estimated number of background events, N_{bkgd} . We need to correct for the finite detector efficiency ϵ . Finally, we also need to allow for the ‘acceptance’ A , which gives the fraction of produced events for which the particle(s) are inside the angular and momentum range that can be detected. Therefore, we modify eqn 13.5 to give

$$\sigma = \frac{N_{\text{obs}} - N_{\text{bkgd}}}{A\epsilon L} \quad (13.6)$$

For these SM measurements, the backgrounds are generally small, so we will postpone a discussion of how to determine them until we consider searches for new physics (see Section 13.5). In principle, the efficiency could be determined purely from a Monte Carlo calculation. However, we can greatly reduce the uncertainty by also using ‘data-driven’ measurements based on the tag-and-probe technique (see Section 13.2). However, the acceptance involves events that are not detectable for a given detector, so we have to rely on Monte Carlo calculations. The Z bosons are identified by detecting two oppositely charged leptons. In the case of electrons and muons, the invariant mass of the pair can be reconstructed and a very clean Z mass peak [22] is observed, as shown in Fig. 13.5 for the case of $Z \rightarrow e^+e^-$.

The $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$ decays can be separated from the background due to dijet events by identifying a well-measured high- p_T electron or a muon and requiring a large value of the missing transverse momentum (see Section 13.2) to identify a neutrino. Nearly all hadrons will be stopped in the calorimeters, so muons can be identified by finding tracks in the muon chambers behind the calorimeters. A very clean peak in the missing-transverse-momentum distribution is observed above the background [63], as shown in Fig. 13.6.

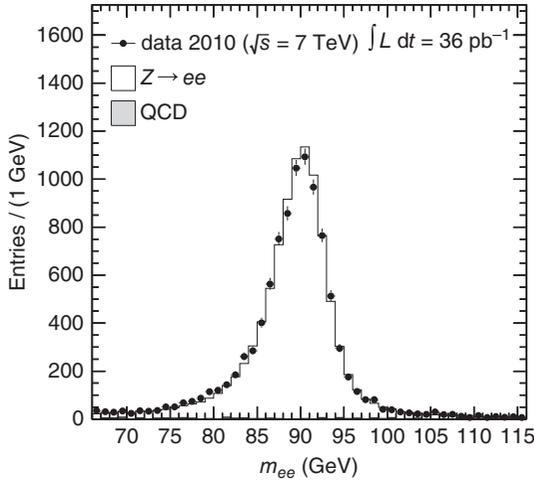


Fig. 13.5 e^+e^- invariant mass spectrum measured by the ATLAS experiment at the LHC [22].

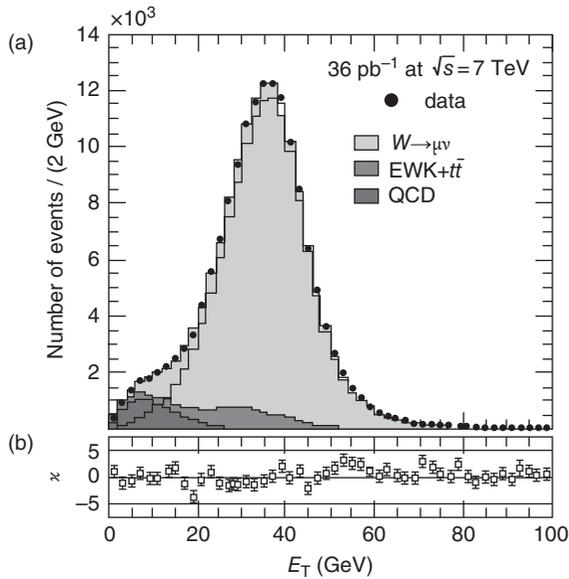


Fig. 13.6 (a) Distribution of missing transverse momentum E_T^{miss} for $W \rightarrow \mu\nu$ measured by the CMS experiment at the LHC [63]. The data are shown with error bars and the backgrounds are shown for other SM sources of genuine muons and for misidentified muons from QCD jets. The dotted histogram gives the fitted signal. (b) Fractional deviation (data - fit)/error.

The cross section for W production at LHC [22] is compared with lower-energy $p\bar{p}$ data and theoretical predictions in Fig. 13.7.

In order to compare theory with data, the calculations are shown for both pp and $p\bar{p}$. The theoretical predictions are in good agreement with all the data. An interesting and more differential test of the SM at LHC is given by measuring the charge asymmetry in W production:

$$A = \frac{\sigma(W^+ \rightarrow l^+\nu) - \sigma(W^- \rightarrow l^-\nu)}{\sigma(W^+ \rightarrow l^+\nu) + \sigma(W^- \rightarrow l^-\nu)} \quad (13.7)$$

At the LHC, W^+ (W^-) will be produced mainly by a valence u (d) quark colliding with a \bar{d} (\bar{u}) quark from the sea (see Chapter 9). Therefore, if

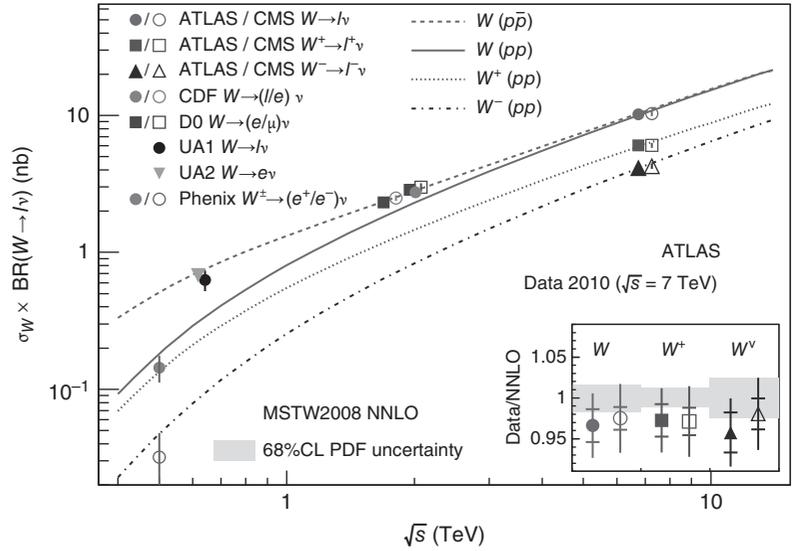


Fig. 13.7 Cross section times branching ratio for W production as a function of CMS energy for $\bar{p}p$ and pp [23].

we assume that the sea distributions are the same for u and d quarks, we should expect

$$A \approx \frac{u(x) - d(x)}{u(x) + d(x)} \quad (13.8)$$

We can determine the p_T of the ν using the missing-transverse-momentum measurement, but we cannot determine the longitudinal momentum of the neutrino. Therefore, as the 4-momentum of the W cannot be uniquely reconstructed, it is convenient to measure the asymmetry as a function of an angular variable. The chosen variable is the pseudorapidity $\eta = -\ln(\tan \frac{1}{2}\theta)$, where θ is the polar angle. In pp interactions, the asymmetry is identical for positive and negative η , so the distribution is symmetric about $\eta = 0$. At $\eta \approx 0$, the parton momentum fractions of the two protons will tend to be similar. So typical x values will be given by $M_W/\sqrt{s} \approx 0.01$ (see Exercise 13.2). For larger values of η , the parton momenta become more unequal and so the asymmetry becomes sensitive to the ratio $u(x)/d(x)$ at larger values of x . From the knowledge of the parton distribution functions (see Chapter 9), we know that $u(x)/d(x)$ increases with x at high x , and, for $x > 0.01$, $u(x)/d(x) > 1$. However, at very large values of η , the effect of the angular distribution of the decay of the W (see Chapter 8) starts to dominate and this generates an asymmetry with the opposite sign. We therefore expect that the asymmetry should be positive and increasing with η over a limited range and then decrease with η in the very forward region. These general features agree with the data from ATLAS, CMS, and LHCb [21], and an NLO QCD prediction (see Chapter 9) is in good agreement with the data, as shown in Fig. 13.8.¹²

¹²These results together with other SM measurements at the LHC can be used to significantly reduce the uncertainty in the parton distribution functions at high values of Q^2 .

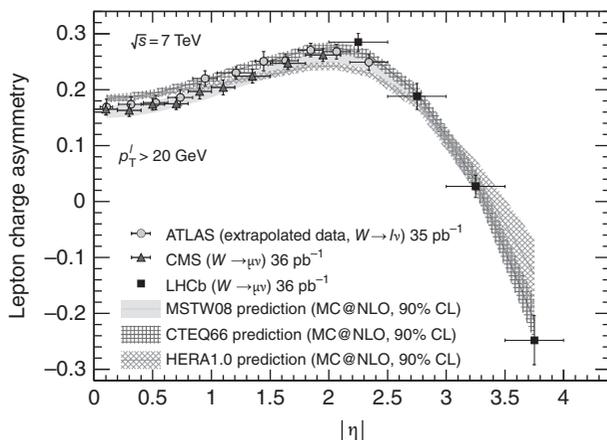


Fig. 13.8 W charge asymmetry as a function of lepton pseudorapidity as measured by ATLAS, CMS, and LHCb at LHC [21].

13.3.1 Top-quark production

Another important test of the SM at LHC is top-quark production. The cross section for $t\bar{t}$ production at a CMS energy of 7 TeV (LHC in 2011) is predicted to be $177.3_{-10.8}^{+10.1}$ pb.¹³ The leptonic decays of the t or \bar{t} will result in events with multijets, lepton(s), and missing transverse momentum from the neutrino(s). These events therefore represent potentially very large backgrounds for new-physics searches based on missing-transverse-momentum signatures. Many measurements of $t\bar{t}$ processes have been made. The cleanest channels are those with both top quarks decaying semileptonically, $t \rightarrow b\nu$. The largest background process is $Z + \text{jets}$. The $Z + \text{jets}$ background can be suppressed by removing events with the same flavour, opposite-sign lepton pairs with an invariant mass consistent with the mass of the Z . The signal events have two neutrinos, which will result in missing transverse momentum (E_T^{miss} , see Section 13.1). The $Z + \text{jets}$ background will only have non-zero E_T^{miss} because of the finite detector resolution. Therefore, this background can be further suppressed by requiring a large value for E_T^{miss} . The signal events will always contain two b quarks, so the purity can be enhanced using ‘ b -tagging’ based on the relatively long lifetime of b hadrons. A likelihood¹⁴ for a given jet to contain a b hadron is constructed based on several variables sensitive to lifetime, including the following:

- The impact parameter (see Chapter 8) divided by its error (called the ‘significance’) for tracks in the jet.
- For each ‘secondary’ vertex, we can determine the decay length from the ‘primary’ vertex. We then define the decay length significance for any secondary vertices associated with the jet as the decay length divided by its error.

A cut on the magnitude of the likelihood is made so as to obtain a b -tagging efficiency of $\sim 80\%$. This technique obviously requires very precise tracking detectors and in particular a very high-precision silicon pixel detector as close to the beam line as possible. The ‘irreducible’

¹³The top cross section for pp at a CMS energy of 7 TeV (LHC in 2011) is a factor of ~ 25 greater than that for $p\bar{p}$ at 2 TeV (Tevatron). At the Tevatron, at the typical value of $x \sim 0.2$, the cross section is dominated by $q\bar{q}$ processes. At the LHC, the smaller x values mean that the cross section is dominated by gg processes.

¹⁴If the probability of a given measurement i under a given hypothesis is p_i , the likelihood is constructed from a set of N measurements as $L = \prod_{i=1}^N p_i$.

background (see Chapter 12 for a discussion of reducible and irreducible backgrounds) is estimated in such a way as to minimize the reliance on Monte Carlo simulations. The general strategy is based on the use of kinematic selections to define ‘control’ regions (in which the events arise mainly from one background source) and ‘signal’ regions. This methodology is explained in Section 13.5.3, where the backgrounds are much more significant. For this measurement, we define a control region in the measured data in which the mass of two same-flavour opposite-sign (SFOS) leptons is compatible with the Z . The Monte Carlo simulation is only used to extrapolate this number to the signal region in which this mass is incompatible with the Z . The distribution of number of jets in this signal region before and after b -tagging [25] is shown in Fig. 13.9. A clean signal is observed above the background even before b -tagging, but the power of b -tagging to greatly enhance the purity of the signal is clearly demonstrated. The measurements of the $t\bar{t}$ cross section at CMS energies of 7 and 8 TeV from ATLAS [33] are shown in Fig. 13.10.

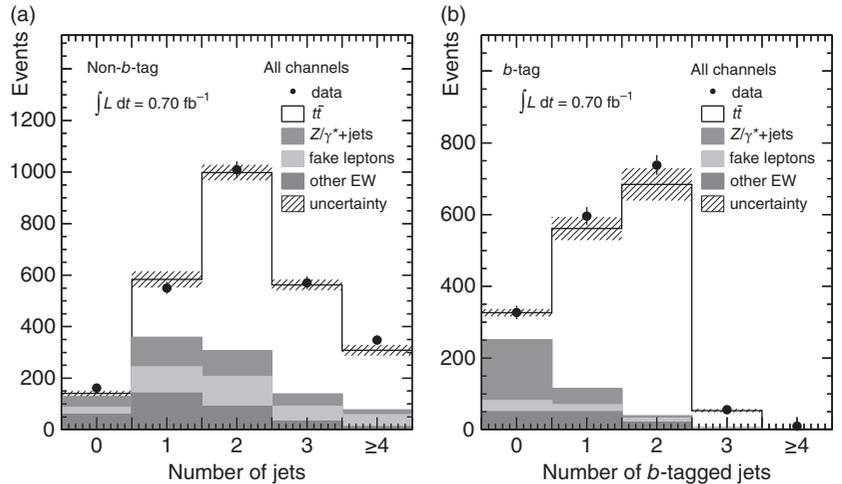


Fig. 13.9 Jet multiplicity distributions for dilepton events in ATLAS for (a) all jets and (b) b -tagged jets [25].

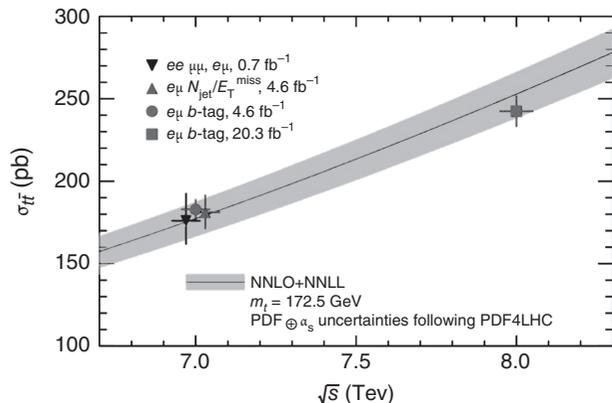


Fig. 13.10 Measurements of the $t\bar{t}$ cross section at 7 and 8 TeV using $e\mu$ b -tag events together with results at 7 TeV using the ee , $\mu\mu$, and $e\mu$ channels measured by ATLAS at the LHC [33].

The good agreement between the data and the SM for these and many other results at the LHC gives us confidence in the reliability of the SM in this energy regime. We can therefore use the SM to reliably predict backgrounds for a wide range of possible new physics processes.

One other very important measurement is the mass of the top quark, because this affects the mass of the W and Higgs bosons via radiative corrections (see Chapter 7). This was used in the past to pin down the mass of the SM Higgs boson, but now that we have observed a Higgs boson, we can combine the three precision mass measurements in a powerful consistency check of the SM. Any significant inconsistency would provide evidence of new physics. The current world average based on the Tevatron measurements by CDF and D0 and the ATLAS and CMS measurements at the LHC gives [19] a value of $m_t = 173.34 \pm 0.76$ GeV. This result is consistent with the SM, but further improvements in the precision will be possible with more data at the LHC.

13.4 Beyond the Standard Model physics

There are several reasons to expect BSM physics to emerge at the TeV scale being explored at the LHC. The first general argument is simply that there are too many free parameters in the SM and a more unified theory should contain fewer. The Higgs mechanism is consistent with current data, but the theory is rather contrived and it is hoped that BSM physics will be able to explain the origin of the spontaneous symmetry breaking required in the Higgs mechanism.

An important argument that points to new physics being manifest at the TeV scale is known as the hierarchy problem in the SM. The mass of the Higgs boson has radiative corrections from the Feynman diagram with a fermion loop (see Fig. 13.11). In this Feynman diagram, we have to consider the propagators for the fermions in the loop. In Chapter 7, we only evaluated Feynman diagrams with boson propagators (e.g. the W^\pm or Z). The Feynman rules for a spin- $\frac{1}{2}$ propagator (see Langacker in Further Reading) of momentum q and mass m give a factor¹⁵

$$f\left(\frac{1}{2}\right) = \frac{i}{\gamma^\mu q_\mu - m} \quad (13.9)$$

Momentum has to be conserved at every vertex in a Feynman diagram, but this allows for an infinite range in the momentum of the virtual fermion. To evaluate this Feynman diagram, we therefore have to integrate over the momentum q of the virtual fermion in the loop. After averaging over the spin states by taking the trace of the fermion propagators [80] of the fermion, we can show that the contribution to the squared mass of the Higgs boson is given by

$$(\Delta m_H)^2 = -C g_f^2 \int_0^\infty d^4 q \operatorname{Tr} \left(\frac{i}{\gamma^\mu q_\mu - m} \frac{i}{\gamma^\mu q_\mu - m} \right) \quad (13.10)$$

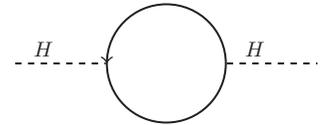


Fig. 13.11 Feynman diagram for the radiative correction to the mass of a Higgs boson from a fermion loop.

¹⁵See Chapter 6 for the definition of the γ_μ matrices.

where g_f is the coupling constant for the fermion–Higgs boson interaction and C contains other numerical constants. Note that eqn 13.10 gives the contribution from one type of fermion, so we need to sum over all fermions. However, the dominant contribution comes from the top quark since it has the largest coupling because it is by far the heaviest fermion. The negative sign arises because there is a fermion loop and this is related to the opposite intrinsic parities of fermions and antifermions. We can write $d^4q = dq_0 d^3\mathbf{q} = dq_0 |\mathbf{q}|^2 d|\mathbf{q}| d\Omega$. Therefore, simply by counting the powers of q , we can see that this integral is ‘quadratically divergent’,¹⁶ i.e. if we introduce a cutoff Λ , the integral will have a term proportional to Λ^2 . The leading term gives

$$(\Delta m_H)^2 = -\frac{g_f^2}{8\pi^2} \Lambda^2 \tag{13.11}$$

If there is no new physics up to some scale Λ , then the mass squared of the Higgs boson will also have radiative corrections of the order of Λ^2 . If Λ is given by the Planck scale or even by a GUT scale,¹⁷ then the natural mass for the Higgs boson would be of the order of 10^{19} GeV or 10^{15} GeV, respectively. We know that the physical mass of the Higgs boson is of the order of 100 GeV, so in the SM this requires counterterms that have to be fine-tuned to 1 part in 10^{15} to ensure the nearly perfect cancellation of the bare mass with the radiative correction. This is called the **hierarchy problem**.

¹⁶We might be worried that we would have similar quadratic divergence for fermion loops between gauge bosons in the SM. However, the SM contains an approximate chiral symmetry that prevents this happening (see Aitchison in Further Reading).

¹⁷The Planck scale, $\sim 10^{19}$ GeV, is the energy scale at which quantum-gravitational effects become important. The GUT scale, $\sim 10^{15}$ GeV, is the scale in a grand unified theory at which the strengths of the electromagnetic and weak interactions become equal.

13.4.1 Supersymmetry

The most popular solution to the hierarchy problem is to invoke supersymmetry (SUSY). This is a symmetry that transforms bosons into fermions and vice versa. This implies that every SM particle must have a superpartner with spin differing by $\frac{1}{2}$. If SUSY is correct, then the Higgs boson mass squared will receive radiative corrections from Feynman diagrams with a scalar loop as shown in Fig. 13.12. The contribution to the square of the μ parameter (see Chapter 12) is given by (see Aitchison in Further Reading)



Fig. 13.12 Feynman diagram for the radiative corrections to the mass of the Higgs boson from a scalar boson loop diagram.

$$\Delta\mu^2 = C\lambda \int_0^\infty \frac{d^4k}{k^2 - m_H^2} \tag{13.12}$$

where C is a numerical constant. As for the calculation of the fermion loop above, we can write $d^4k = dk_0 d^3\mathbf{k} = dk_0 |\mathbf{k}|^2 d|\mathbf{k}| d\Omega$, which contains the fourth power of k , and hence this integral is also ‘quadratically divergent’. This means that if we introduce a cut-off parameter Λ to the integral, we find (C' is another numerical constant)

$$\Delta\mu^2 = C'\lambda\Lambda^2 \tag{13.13}$$

The radiative correction to the μ^2 parameter then introduces a correction to the mass squared of the Higgs boson of

$$(\Delta m_H)^2 = \frac{g_s}{16\pi^2} \Lambda^2 \tag{13.14}$$

where g_s is the coupling constant for the scalar interaction.

Comparing eqns 13.11 and 13.14, we see that we will have perfect cancellation of the unwanted quadratic divergence if $g_f^2 = g_s$ and we have two scalar partners for every SM fermion. This happens naturally in SUSY, which is an extension of the symmetries of the SM relating bosons and fermions. In SUSY, the two scalar particles for each SM fermion arise from the left- and right-handed fermions. The equality of the coupling constants is guaranteed by the symmetry.

Unbroken SUSY therefore solves the hierarchy problem and introduces no new parameters into the theory. However, SUSY is manifestly broken, because we have not yet discovered any of the SUSY partners of the SM particles, so their masses must be greater.¹⁸ However, if the masses are very much heavier than the SM partners, then some fine tuning would be required to prevent the Higgs boson mass becoming too large. Therefore, ‘naturalness’ arguments suggest that the masses of the SUSY partners should be of the same order of magnitude as the Higgs vacuum expectation value (246 GeV).

There are many different options for how SUSY is broken.¹⁹ The masses of the superpartners are not protected by the gauge symmetries that force the masses of fermions and gauge bosons to be zero before spontaneous symmetry breaking is considered. In this sense, it is not surprising that the masses of SUSY particles are larger than those of their SM particles.

The particle content of the SUSY extension to the first generation of the SM is given in Table 13.1. SUSY particles are labelled with a ‘tilde’ ($\tilde{}$) above their name; for example, the SUSY partner of the electron is the selectron \tilde{e} . SUSY partners of quarks and leptons are called squarks and sleptons, respectively. SUSY partners of neutrinos, gauge bosons, and Higgs bosons are called neutralinos, gauginos, and Higgsinos.

After SUSY breaking, there will in general be mixing of all states with the same quantum numbers. For example the ‘left’ and ‘right’ t -quark states will mix to form the physical states called \tilde{t}_1 and \tilde{t}_2 . This mass splitting can be very large, so it is possible that the lightest squark belongs to the third generation. The Higgs sector in SUSY is extended compared with the SM because two complex Higgs doublets are required to give masses to the up-type and down-type quarks. Each complex doublet consists of four fields. After spontaneous symmetry breaking, three of the eight fields are ‘swallowed up’ to give mass to the W^\pm and Z^0 bosons. This leaves five physical spin-0 fields: two charged fields (H^\pm) and three neutral fields (a pseudoscalar A and two neutral scalars h and H). There are also SUSY partners for these states. There is then mixing between the SUSY partners of the neutral gauge bosons $\tilde{\gamma}$, \tilde{Z}^0 and the neutral Higgsino states \tilde{h} , \tilde{H} to give the physical states $\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$, where the states are labelled in order of increasing mass. Similarly, the \tilde{W}^\pm states mix with the \tilde{H}^\pm states to give the charginos $\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$.

Although there is currently no direct evidence for SUSY, there are some suggestive hints that it might be a valid low-energy theory (i.e. at the electroweak scale ~ 100 GeV):

- SUSY helps with the unification of the electroweak and strong coupling constants. If we assume that there should be some grand

¹⁸There is clearly no scalar charged particle with the same mass as the electron (511 keV).

¹⁹SUSY breaking must be done in such a way as to preserve the gauge symmetries in the SM, so that the SM particles only acquire mass through spontaneous symmetry breaking.

SM particles	Spin	SUSY partners	Spin
u_L, u_R	$\frac{1}{2}$	\tilde{u}_L, \tilde{u}_R	0
d_L, d_R	$\frac{1}{2}$	\tilde{d}_L, \tilde{d}_R	0
e_L, e_R	$\frac{1}{2}$	\tilde{e}_L, \tilde{e}_R	0
ν_e	$\frac{1}{2}$	$\tilde{\nu}_e$	0
γ	1	$\tilde{\gamma}$	$\frac{1}{2}$
W^\pm	1	\tilde{W}^\pm	$\frac{1}{2}$
Z^0	1	\tilde{Z}^0	$\frac{1}{2}$
g	1	\tilde{g}	$\frac{1}{2}$

Table 13.1 Example of SUSY particles for the first generation of quarks and leptons and for the gauge bosons.

²⁰The renormalization group equations are derived from the underlying group theory and allow the determination of the change of coupling ‘constants’ with scale. See the discussion in Chapter 9 for a simple explanation in the context of QED and QCD.

unification of the electroweak and strong couplings at some high energy, then we can use the ‘low-energy’ data and the renormalization group equations²⁰ to see if this happens. In the context of the SM, the coupling constants do not quite coincide at any scale. However, if one assumes SUSY with a mass scale of the order of a TeV, then the coupling constants do coincide as shown in Fig. 13.13.

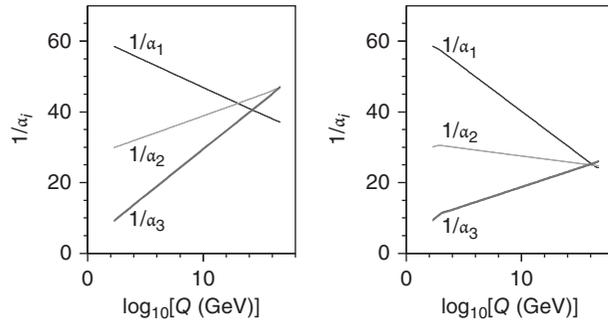


Fig. 13.13 Evolution of the coupling constants in the SM and in SUSY [70].

- In SUSY models, we can ensure that the lightest supersymmetric particle (LSP) will be stable (see Section 13.4.2) and will therefore provide a natural candidate for dark matter. This requires SUSY particles at a mass scale of $\sim \text{TeV}$.
- In the SM, spontaneous symmetry breaking has to be put in by hand. The mass terms in the Lagrangian evolve with scale ('run'). If the mass-squared term for the Higgs potential is positive at some high (unification) scale, then the large top-quark Yukawa coupling will drive this term negative (see Chapter 12), thus giving some explanation for the existence of spontaneous symmetry breaking at low energy. This argument also requires that the mass scale of the SUSY partners is $\sim \text{TeV}$ (see Martin in Further Reading).
- In the SM, the Higgs boson mass is unconstrained. In the minimal SUSY SM (MSSM) at 'tree level',²¹ the Higgs boson mass should be given by [83]

$$m_H^2 = M_Z^2 \cos^2 2\beta \quad (13.15)$$

where β is the ratio of the vacuum expectation value of the Higgs boson coupling to up-quark and down-quark flavours. Therefore, at tree level, the Higgs boson is constrained to be lighter than the Z . However, there are radiative corrections to the Higgs boson mass and an upper limit is given by [83]

$$\Delta m_H^2 = \frac{3g^2 m_t^4}{8\pi^2 M_W^2} \left[\ln \left(\frac{M_s^2}{m_t^2} \right) + \frac{X^2}{M_s^2} \left(1 - \frac{X_t^2}{12M_s^2} \right) \right] \quad (13.16)$$

where g is the weak coupling constant, M_s is the geometric mean of the stop quark masses ($M_s^2 = m_{\tilde{t}_1} m_{\tilde{t}_2}$), and X_t is a stop mixing parameter. The MSSM still predicts a relatively low value $m_H < 140 \text{ GeV}$, which is consistent with the recent results from Higgs boson searches at the LHC, which give $m_H \sim 125 \text{ GeV}$.

²¹By tree level we mean a calculation using only the lowest-order Feynman diagrams.

13.4.2 R -parity

One potentially fatal problem with SUSY is that in general it will allow quarks to convert to leptons and thus allow Feynman diagrams that would mediate very rapid proton decay as shown in Fig. 13.14.

The simplest solution to this problem is to invoke a new multiplicative parity, called R -parity, given by $R_p = (-1)^{3(B-L)+2s}$, where B is the baryon number, L the lepton number, and s the spin. All the SM particles have $R_p = 1$ and all the SUSY partners have $R_p = -1$. This has profound consequences:

- The lightest SUSY particle (LSP) is absolutely stable, since any decay to SM particles would violate R -parity. This has the attractive feature that the LSP is a natural candidate for dark matter (see Section 13.7).

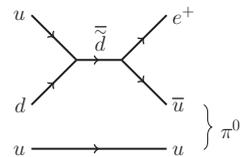


Fig. 13.14 Feynman diagram mediating proton decay in SUSY, $p \rightarrow e^+ \pi^0$.

- Heavier SUSY particles will decay to an odd number (usually one) of SUSY particles (and SM particles).
- SUSY particles will always be produced in pairs in collider experiments.
- The LSP will have very small cross sections for interactions with ordinary matter because any interaction will be suppressed by the high SUSY mass scale. Therefore, from the experimental point of view, LSPs will behave like neutrinos by not interacting in detectors. It is assumed that the LSP is neutral, because otherwise there would be a measurable relic density in nuclei.

Even in the context of R -parity-conserving SUSY models, there are far too many parameters to allow a model-independent survey of the full parameter space. One way to get round this problem is to assume a simplified model in which some of the key parameters of the MSSM are set by hand.

13.4.3 Other BSM theories

Another solution to the hierarchy problem is to replace the fundamental Higgs scalar with a composite object. Theories based on this idea are called ‘technicolour’, but are not discussed here since it is difficult to construct such a theory that is compatible with all the precision electroweak data. Technicolour theories are also disfavoured by the LHC observation of a boson that is compatible with the properties of the SM Higgs boson.

Another interesting option is to assume that there are ‘large’ extra dimensions. In these models, the SM particles are restricted to a four-dimensional ‘brane’ but gravity can propagate in the higher-dimensional ‘bulk’. Gravity then appears weak in our four-dimensional world because the gravitons can propagate into the bulk. At distances shorter than the characteristic length of the theory, there would be deviations from the inverse square law for gravity. However, experiments to detect such an effect have not found any deviation from the inverse square law. In these models, the fundamental Planck scale can be as low as \sim TeV. This means that the cut-off in the divergent integral for the Higgs boson mass is \sim TeV, thereby avoiding the hierarchy problem.

There are many versions of these models, but one generic feature is that at TeV scales particles can leak from the brane into the bulk and thus cause an apparent violation of momentum conservation in four dimensions. Therefore, from an experimental point of view, we would expect events with large missing transverse momentum. More speculatively, some of these models predict the production of micro black holes at the LHC with significant cross sections. The non-observation of micro black holes at the LHC therefore provides severe constraints on these theories.

13.5 Experimental searches for BSM physics at the LHC

There are a bewilderingly large number of possible BSM theories that might be experimentally accessible. It is therefore essential that some searches are geared to finding unexpected signals while others are optimized for some particular model, such as SUSY. This section merely aims to give the flavour of these searches.

13.5.1 Jet production

The large QCD jet production cross sections at LHC provide access to significant numbers of events at very high mass. These can be used to search for resonances in the mass spectra as would be expected from models like excited quarks as well as new contact interactions. A particular sensitive way to search for new physics is to use the angular distribution in the dijet CMS, because many experimental systematic effects (such as the jet energy scale) are thereby greatly reduced. The angular distribution is studied as a function of the variable

$$\chi = \frac{1 + \cos \theta^*}{1 - \cos \theta^*} \quad (13.17)$$

where θ^* is the polar angle of one of the jets in the dijet CMS system. The advantage of the χ variable is that the forward-peaked angular distribution for θ^* is transformed into a more uniform distribution as a function of χ . New physics such as a contact interaction will tend to produce a more isotropic distribution in $\cos \theta^*$ and therefore a peak at small values of χ . The distribution of χ for a range of dijet invariant masses [32] is shown in Fig. 13.15. No significant deviation from the SM is seen and limits on new physics can be placed.²²

²²In this case, limits were placed on models with extra dimensions, but the data can be used to place limits on other BSM physics.

13.5.2 Lepton pair resonances

Many BSM theories predict extra U(1) or SU(2) symmetries, which would result in heavier versions of W and Z bosons, called W' and Z' . The signature for a W' would be a Jacobian peak in the transverse mass (see Chapter 8, eqn 8.22) above M_W . The signature for a Z' would be a narrow peak in the l^+l^- invariant mass distribution above M_Z . These would give clear signals above the SM backgrounds, but so far only upper limits have been reported; see for example Fig. 13.16 [27].

13.5.3 SUSY searches

If we assume R -parity-conserving SUSY, then SUSY particles will be produced in pairs, such as $\tilde{q}\tilde{q}$. The cross sections can be calculated at the parton level, since all the couplings are the same as in the SM. Using the measured parton distribution functions (see Chapter 9), we can then

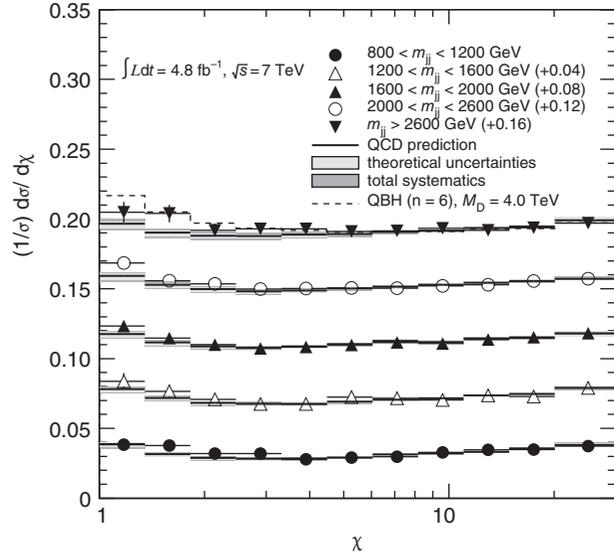


Fig. 13.15 Distribution of χ for slices of dijet invariant mass, measured by ATLAS and compared with QCD predictions (shaded bands) and a BSM theory with extra dimensions (dashed line) [32].

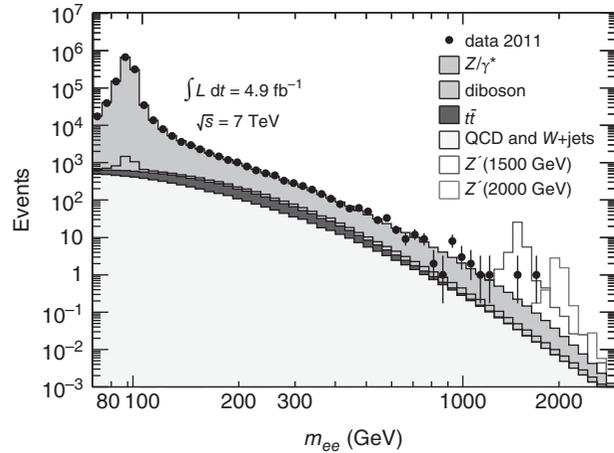


Fig. 13.16 Distribution of dilepton invariant mass, measured by ATLAS and compared with SM expectations and models with different Z' masses [27].

calculate the cross section for SUSY pair production in pp collisions. The results are shown in Fig. 13.17 as a function of the average mass of the pair of sparticles [119]. There is a relatively large cross section for strongly interacting squarks and gluinos, but the cross sections for particles produced via electroweak interactions are much smaller. The cross section for a given type of particle decreases rapidly with the mass of the particle.

While the production cross sections are well defined, the decay modes for a given sparticle depend on which modes are kinematically allowed. They also depend on the model parameters that are used for SUSY breaking. However if we assume R -parity conserving SUSY, then the end result of the decays of two sparticles will be two LSPs. The LSPs

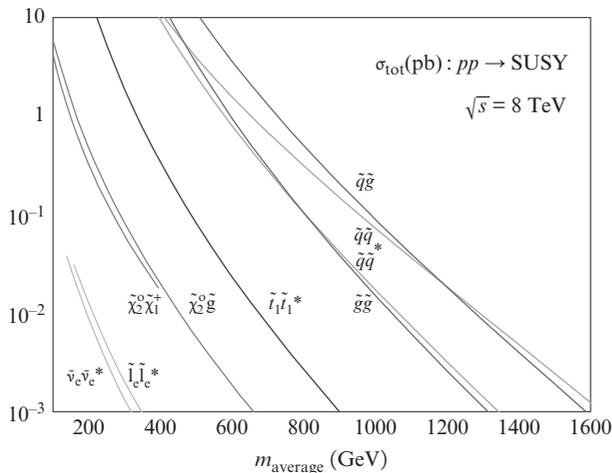


Fig. 13.17 Predicted cross section for the production of pairs of SUSY particles for pp collisions at a CMS energy $\sqrt{s} = 8$ TeV. From [119].

will not be detected directly in a detector. Therefore, a large value of E_T^{miss} will be used as a signature for SUSY searches. In order to establish evidence for a possible SUSY signal, all possible sources of background must be understood. After removing instrumental backgrounds such as ‘beam halo’ interactions²³ and fake E_T^{miss} generated by the finite detector resolution, the main backgrounds are SM processes that have genuine E_T^{miss} from final-state neutrinos. Some of the main SM backgrounds are

- (1) Z + jets with $Z \rightarrow \nu\bar{\nu}$;
- (2) W + jets with $W \rightarrow l\nu$;
- (3) $t\bar{t}$ with at least one top quark decaying semileptonically to produce neutrinos.

²³‘Beam halo’ refers to beam particles that have escaped from the beam pipe.

The first reaction is an example of an irreducible background in that even with a perfect detector one could not separate it from signal on an event-by-event basis. Therefore, it is essential to have a very reliable prediction of the rate for this process. This can be done in a largely data-driven way, thus minimizing theoretical uncertainties: the rate for Z + jets with $Z \rightarrow l^+l^-$ can be measured directly, and if the l^+l^- are removed, this provides a simulation of $Z \rightarrow \nu\bar{\nu}$. The ratio of branching ratios of Z to leptons and neutrinos is very well known from LEP. The effect of finite efficiencies for triggering and identifying charged leptons can be measured in a data-driven way using $Z \rightarrow l^+l^-$ events. A Monte Carlo calculation is only required to allow for the finite detector acceptance of the charged leptons (the acceptance for neutrinos is 100%). The other reactions generate ‘reducible’ backgrounds that could be completely rejected with an ideal detector because they result in charged leptons, whereas some of the SUSY signal will not contain any charged leptons. Data-driven techniques are required to estimate the effects of the finite detector efficiency.

The kinematics of SUSY signals and SM backgrounds are very different and this is exploited in SUSY searches. SUSY events will typically contain more high- p_T jets than the SM backgrounds, and SUSY events will tend to be at larger values of E_T^{miss} than those from the SM. At low values of E_T^{miss} , the distribution will be dominated by the SM backgrounds but any significant excess at high values of E_T^{miss} would be evidence for SUSY.

A general approach to estimating the SM backgrounds in searches for BSM physics is based on using Monte Carlo simulations to

- (1) Define suitable signal regions (SR) in order to optimize the discovery potential for a particular range of parameters of a model. That is, we want to maximize the acceptance for BSM physics while simultaneously minimizing the number of SM background events that would be expected.
- (2) Define appropriate control regions (CR) that isolate one particular SM process. In such a region, we would expect this process to be dominant and we would not expect any significant ‘contamination’ from BSM physics.

We can then measure the number of events in the control region in real data ($N_{\text{data}}^{\text{CR}}$) and Monte Carlo simulated data ($N_{\text{MC}}^{\text{CR}}$). Finally, we need to use the Monte Carlo simulation to calculate the ‘transfer factor’ (TF), defined by

$$\text{TF} = \frac{N_{\text{MC}}^{\text{SR}}}{N_{\text{MC}}^{\text{CR}}} \quad (13.18)$$

²⁴This is a simplified discussion, since it assumes that each control region is only populated by events from one SM process. Nevertheless, it should give a clear idea of the approach used. Note that it does not entirely eliminate the reliance on Monte Carlo simulations, but it only requires the Monte Carlo predictions of ratios, so the systematic uncertainties are greatly reduced compared with relying on Monte Carlo to predict the absolute number of SM events in the SR.

The predicted background for a given process²⁴ is then given by

$$N_{\text{predicted}} = \text{TF} \times N_{\text{data}}^{\text{CR}} \quad (13.19)$$

While this approach works well for SM processes involving W, Z , or top-quark production, it cannot be used to predict QCD-induced backgrounds such as ‘jets’ that are misidentified electrons or ‘fake’ E_T^{miss} arising from the finite detector resolution. In general, the QCD cross sections are so much larger than those for BSM physics that a data-driven method must be used. There are several ways of doing this. One option is to perform a ‘template’ fit. For example, to estimate the background from ‘fake’ electrons (i.e. QCD jets wrongly identified as electrons):

- A pure QCD sample is defined by reversing electron selection criteria, and this is used to measure the shape of the background distribution.
- The Monte Carlo simulations are used to predict the shape of the sum of the other SM processes, excluding QCD multijets.

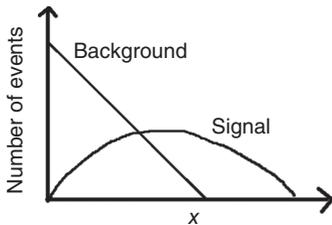


Fig. 13.18 Distributions for signal and background. The x axis is a proxy for a suitable variable like E_T^{miss} .

This method works if the shapes of the two distributions are sufficiently different, as indicated in the sketch in Fig. 13.18. The data distribution

is then fitted to the sum of the two distributions, and the fit can then be used to determine the number of background events in any SR.

As discussed in Section 13.4.1, the number of free parameters in the minimal SUSY model (MSSM) are so large that it is not feasible to search the full parameter space. One example of a search [36] for squarks was performed with the assumption²⁵ that the gluinos are much heavier than the squarks and that the squarks decay with 100% branching ratio as $\tilde{q} \rightarrow q\chi_1^0$. The signature for the strong-interaction production of pairs of squarks is therefore (at least) two high- p_T jets plus large E_T^{miss} . Another SUSY search targets very large jet multiplicities that can arise from cascade decays of SUSY particles. The dominant background source for this analysis was from mismeasured values of E_T^{miss} . The experimental uncertainty on the measurement of E_T^{miss} is found to scale as $E_T^{\text{miss}} \propto \sqrt{H_T}$, where H_T is the sum of the transverse momenta of the jets. Therefore, the signal (control) region is defined by events with large (small) values of $E_T^{\text{miss}}/\sqrt{H_T}$. The distribution of $E_T^{\text{miss}}/\sqrt{H_T}$ was measured for a lower-jet-multiplicity sample and this was then used to scale the number of background events in the control region to the signal region. The resulting distribution [31] of $E_T^{\text{miss}}/\sqrt{H_T}$ is shown in Fig. 13.19. The data are in good agreement with the background calculations, and powerful limits on the masses of the squarks and gluinos can be obtained as shown in Fig. 13.20. The curves show that for a particular simplified SUSY model, a clear excess would have been expected at large values of $E_T^{\text{miss}}/\sqrt{H_T}$.

The hierarchy argument suggests that the mass of the stop squark should be relatively low (see eqn 13.16) but the masses of the other

²⁵Other analyses make different model assumptions.

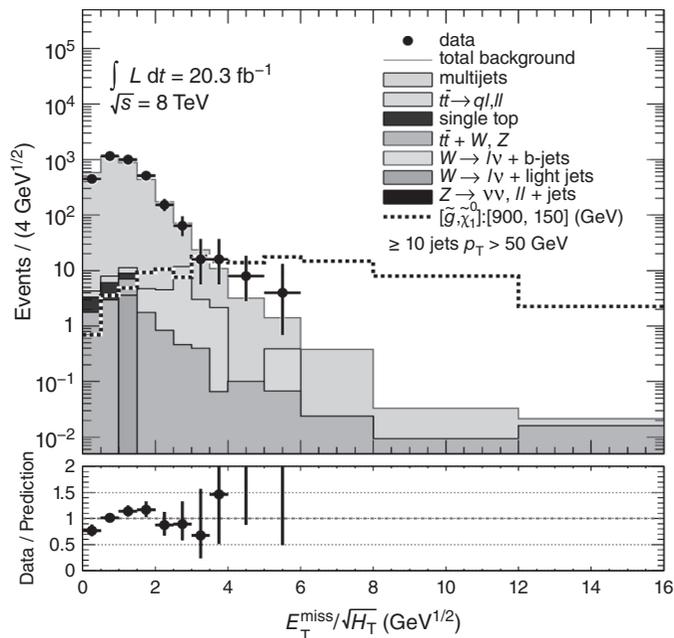


Fig. 13.19 Distribution of $E_T^{\text{miss}}/\sqrt{H_T}$ measured by ATLAS in a signal region with the number of jets ≥ 10 . The SM backgrounds and the expected signal for SUSY signal for a simplified model with masses given by $m(\tilde{g}) = 900$ GeV and $m(\chi_1^0) = 150$ GeV are shown [31].

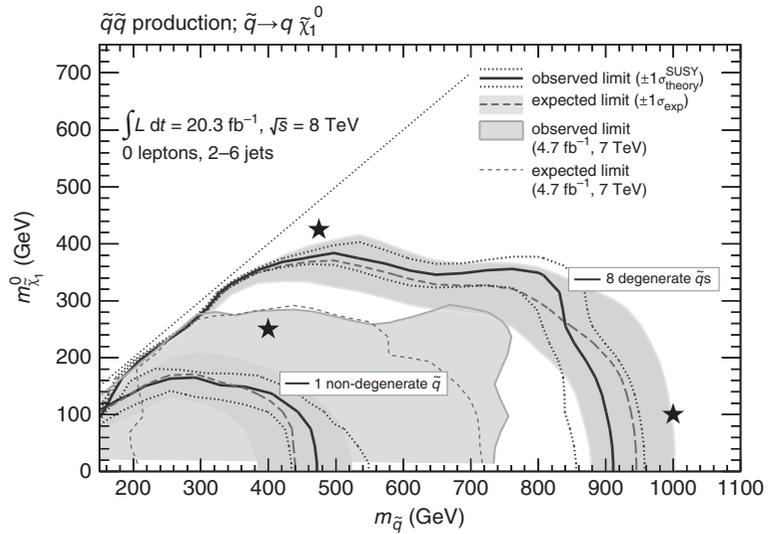


Fig. 13.20 Limits in the plane of the mass of the $\tilde{\chi}_1^0$ versus the mass of the lightest \tilde{q} , obtained in simplified models with the assumption of one non-degenerate squark or eight degenerate squarks [31]. The areas below the solid curves are excluded at 95% confidence level.

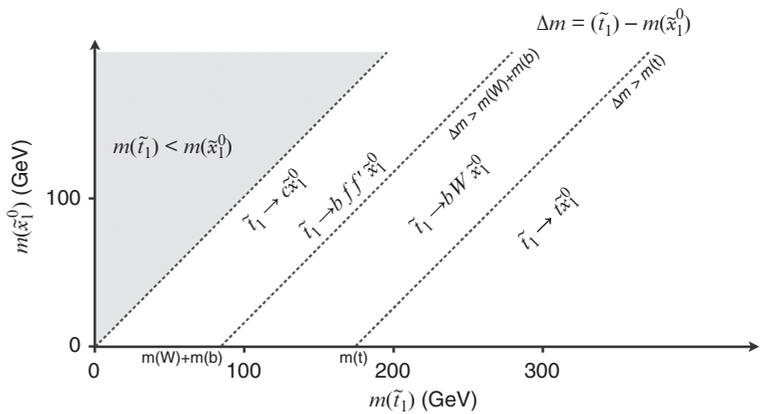


Fig. 13.21 Kinematically allowed decay modes of the \tilde{t} for different regions in the mass plane $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$ [34].

squarks could be much higher without violating ‘naturalness’. Therefore, the search for stop squarks is particularly important. In general, the SUSY signal for $t\tilde{t}$ production will suffer from SM backgrounds from $t\bar{t}$ production. The kinematically allowed decay modes for the stop depend on the masses of the stop and the neutralino, as illustrated in Fig. 13.21. Several searches for the stop have been performed. One particularly powerful search used final states with one-lepton and multijets [34]. The analysis used several signal regions to target the kinematics for the different regions in the mass plane $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$. The results were consistent with the SM, and the resulting limits are shown in Fig. 13.22.

In some SUSY models, the squarks and gluinos are very heavy and the lightest SUSY particles are the charginos and neutralinos. Therefore, another interesting way of searching for SUSY is to select events with charged leptons and large values of E_T^{miss} . There are many possible SUSY

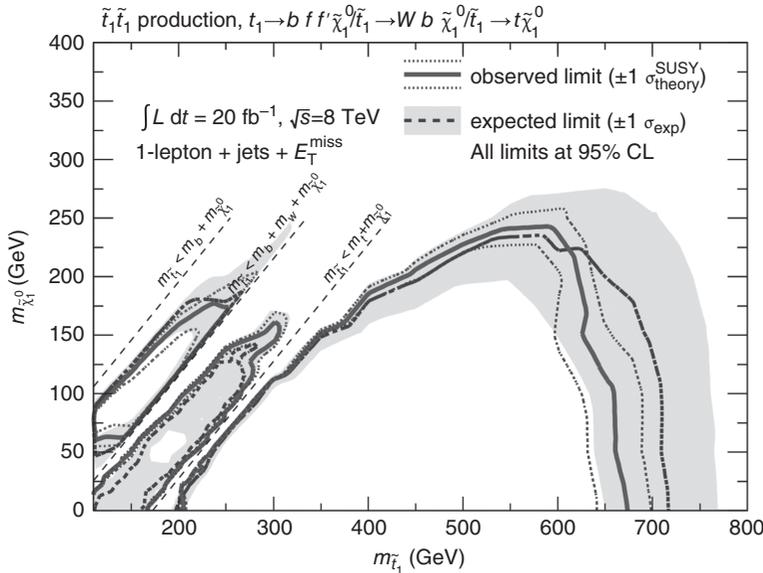


Fig. 13.22 Limits in the plane of the mass of the χ_1^0 versus the mass of the \tilde{t} [34]. The areas below the solid curves are excluded at 95% confidence level.

production and decay chains that would lead to these final states. One analysis [35] considered the electroweak production of $\chi_1^\pm \chi_2^0$, followed by either the decay chain $\chi_1^\pm \rightarrow \tilde{\nu} l$ and $\tilde{\nu} \rightarrow \nu \chi_1^0$ or $\chi_1^\pm \rightarrow \tilde{l} \nu$ and $\tilde{l} \rightarrow l \chi_1^0$, together with similar decay modes for the χ_2^0 (see Fig. 13.23). The results shown in Fig. 13.24 represent a large improvement on the limits from LEP2.²⁶ Many other electroweak processes can be considered in the search for charginos and neutralinos. For example, the Feynman diagram for the electroweak production of $\chi_1^0 \chi_1^+$ is shown in Fig. 13.25. In many models, the decays of the χ_1^+ result in a charged lepton. This results in a distinctive experimental signature of a charged lepton and a large value of E_T^{miss} .

²⁶At LEP2, the limits on charged-particle production are to a good approximation given by the beam energy, since charged particles are produced in pairs.

13.5.4 Summary of searches for new physics

The limits on the masses of SUSY particles and other exotic physics have been greatly extended by the LHC, but as of 2015 there is no clear evidence for any new physics at the LHC. However, the mass reach will be greatly extended by the increase in energy to 13 TeV and the planned increase in luminosity up to and beyond the nominal value of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, so there are prospects for exciting discoveries in the next few years

13.6 Linear collider

The observation at LHC of a low-mass Higgs-boson-like particle gives a strong motivation for studying its properties in more detail in an e^+e^-

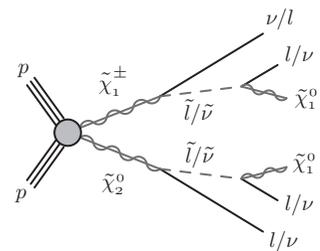


Fig. 13.23 Electroweak SUSY production processes for $\chi_1^\pm \chi_2^0$, with decays via sleptons or sneutrinos [35].

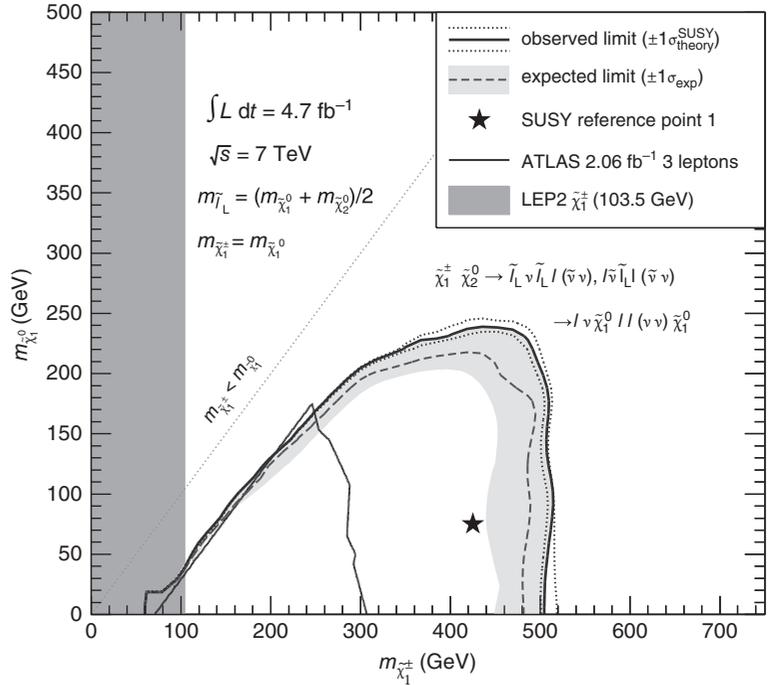


Fig. 13.24 Mass limits in the plane of the masses of the χ_1^0 and χ_1^\pm in a simplified SUSY model [35].

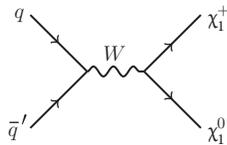


Fig. 13.25 A Feynman diagram for electroweak chargino-neutralino production.

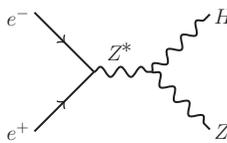


Fig. 13.26 Production of Higgs bosons in e^+e^- annihilation.

collider. In an e^+e^- collider, a Higgs boson would be produced by the ‘Higgsstrahlung’ process (see Fig. 13.26) $e^+e^- \rightarrow HZ$. The minimum CMS energy for this process is $m_Z + m_H$, but a useful rule of thumb is that to get a sufficiently large cross section requires an energy of $\sim m_H + 100$ GeV. Given the mass of the boson observed at LHC (see Chapter 12), the minimum energy of an e^+e^- collider to be operated as a ‘Higgs factory’ would be ~ 230 GeV. In such a process, the events could be tagged by detecting the Z decay products, so that all decay modes of the Higgs boson could be studied precisely to determine their branching ratios. In the SM, all the branching ratios of the Higgs boson can be calculated once its mass has been determined, and therefore this would provide stringent tests of whether the properties of the observed particle are consistent with those expected in the SM. Using polarized electron beams would also allow a determination of the spin and parity of the Higgs boson (the LHC measurements strongly support the SM assignment of a spin/parity of 0^+).

However, to confirm that the particle really was a Higgs boson, we would need to confirm that the self-coupling of the Higgs boson exists with the expected strength given by the SM. This could be studied with higher luminosity at the LHC or at an e^+e^- collider if it had sufficiently high CMS energy. If the CMS energy of a linear collider were larger than twice the top mass, many precise measurements would be possible. The mass of the top quark could be measured with a precision an order of magnitude better than achievable at LHC. By comparing

the measurements of the mass of the W , the top quark, and the Higgs boson, a precision test of the radiative corrections in the SM would be possible. The interest here is that any deviations from the SM prediction would give indications of new physics at higher mass scales, through loop diagrams. Precision measurements of the top-quark axial and vector couplings could be made using the electron polarization and again any deviations from the SM predictions would give sensitivity to new physics.

If the CMS energy were large enough, an e^+e^- collider could be used to study SUSY particles. This would require that the beam energy be larger than the lightest SUSY charged particle. This might require a much higher-energy machine than would be required for the study of the Higgs boson and top quark. Such a machine would have to be a linear collider, because a circular machine would have unacceptable synchrotron radiation losses or require an unrealistically large radius. R&D for a linear collider is based around the development of high-gradient superconducting radiofrequency (RF) cavities. Currently, gradients of $\sim 35 \text{ MeV m}^{-1}$ have been achieved. In order to achieve useful interaction rates, very high luminosity would be required. As a linear collider is a single-pass machine (unlike circular machines), this will require very small beam sizes. This makes many demands on the machine:

- high-quality, low-emittance beams of both electrons and positrons;
- powerful final focus to reduce the beam sizes in the vertical dimension to $\sim 1 \text{ nm}$.
- precise alignment of the electron and positron beams and fast-feedback systems to ensure that the beams collide.

Considerable progress has been made in these issues, but more remains to be done before such a machine could be built.

Looking further into the future, if a higher-energy e^+e^- collider were to be built, it would probably require the more exotic technology being developed for the CLIC (Compact Linear Collider). This is aimed at a collider with a CMS energy around 3 TeV. To keep the length affordable, accelerating gradients of $\sim 100 \text{ MeV m}^{-1}$ would be required, which is too large for the superconducting RF technology (see Chapter 3) being developed for lower-energy linear colliders. Such gradients could be achieved using a two-beam-acceleration concept in which the RF power is generated by a very high-current ($I \sim 100 \text{ A}$) electron beam (drive beam) running parallel to the main beam. This drive beam is decelerated and the generated RF power is transferred to the main beam.

13.7 Dark matter

There are strong indications from astrophysics that the universe contains a large amount of non-baryonic matter. This matter only has weak and gravitational interactions and cannot emit electromagnetic radiation,²⁷ hence the name ‘dark matter’. The evidence for this dark matter comes

²⁷There are models in which dark matter only has gravitational interactions, but they are not discussed here as they are almost impossible to test directly experimentally.

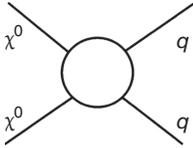


Fig. 13.27 General picture of quark–WIMP scattering. The WIMPs are labelled as χ^0 .

from its gravitational interaction with the luminous normal matter and this is briefly reviewed in Section 13.7.1. If such dark matter exists in the form of weakly interacting massive particles (WIMPs), there are three different approaches to studying them in laboratory or astroparticle experiments, which can be explained in terms of the cartoon shown in Fig. 13.27.

- (1) **Direct detection:** In this approach, we use the interactions of a WIMP with a quark in a nucleus and try to detect the resulting nuclear recoil. This is discussed in Section 13.7.2.
- (2) **Accelerator detection:** If we collide quarks (or other partons) at sufficiently high energy, we can pair-produce WIMPs. As WIMPs will not interact in the detector, this process is observable in events in which one of the initial-state quarks radiates a gluon or a W or Z , since this then results in events with large values of E_T^{miss} . These events can be searched for in the same way as SUSY (see Section 13.5.3).
- (3) **WIMP annihilation:** If WIMPs accumulate under gravity, there will be significant rates of annihilation. The search for these reactions is discussed in Section 13.7.3.

This qualitative comparison between the accelerator searches and direct detection can be made more quantitative using effective field theories (EFTs) to describe the interactions between quarks and WIMPs. An EFT provides a low-energy approximation to a fuller but not yet known theory. The full theory would give accurate predictions at high energies, whereas the EFT would be expected to fail at high energies. If the separation between the two energy scales is large enough, the EFT may give reliable predictions at low energy.²⁸

13.7.1 Astrophysical evidence for dark matter

There are several aspects of the astrophysical evidence for dark matter. One approach uses galactic rotation curves. The velocities of stars can be measured from their Doppler shifts and these can be compared with the velocities calculated assuming Newtonian gravitation. The latter calculation assumes that the distribution of gravitational matter follows that of the luminous matter. In this case, using Newtonian mechanics and the inverse square law of gravity,²⁹ one expects to find the rotation velocity $v(r)$ scaling with distance r from the galactic centre as $v(r) \propto 1/\sqrt{r}$ for values of r larger than the bulk of the luminous matter of the galaxy. However, the rotation curves are usually much flatter at large r (see Fig. 13.28 for an example of a galactic rotation curve [47]). If one assumes that the inverse square law of gravity is correct, then there must be a halo of non-luminous matter, called ‘dark matter’. To produce a flat rotation curve, the dark matter mass contained within a radius r must scale like $M(r) \propto r$ or the density must scale like $\rho(r) \propto 1/r^2$ (at

²⁸A classic example of an EFT is the Fermi theory of weak interactions, which ignores the effects of the W propagator and treats the interaction as a four-particle interaction. The Fermi theory works very well at low energies (~ 10 MeV) because of the very high mass of the W boson.

²⁹An alternative attempt to explain this and other effects is to assume that the inverse square law needs to be modified at large values of r .

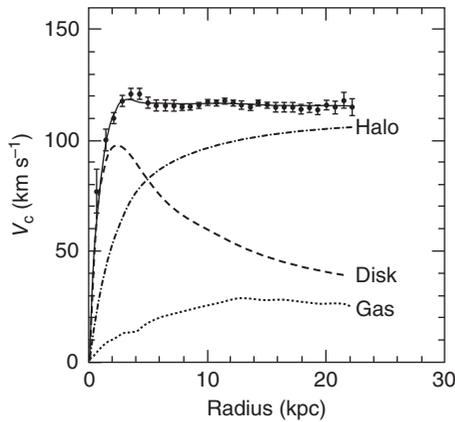


Fig. 13.28 Rotation curve for the galaxy NGC 6503 [47].

very large values of r , the density must fall off faster in order for the galactic mass to be finite).

Additional evidence comes from the motion of galaxies in clusters of galaxies. A striking piece of evidence for dark matter comes from the observation of the Bullet Cluster, in which two clusters have collided with one another. The distribution of gravitational mass can be inferred from gravitational lensing (see Perkins in Further Reading) and is found to be very different to the distribution of baryonic mass as determined from the observed light. The baryonic matter interacted via electromagnetic interactions and shows evidence for shock waves. Therefore, the interactions between the dark matter and the baryonic matter must be very much weaker than this.

From global fits to these and other data, the dark matter content of the universe is much greater than that of ordinary baryonic matter. Direct dark matter searches are discussed in Section 13.7.2 and the search for dark matter annihilation is discussed in Section 13.7.3. The accelerator (LHC) search is not discussed further here, since the methodology is similar to that of the search for events with large values of E_T^{miss} that has been discussed in the context of SUSY.

13.7.2 Direct dark matter detection

At first sight, it might seem surprising that it is difficult to detect dark matter if there is much more dark matter than ordinary matter. The problem is that the dark matter particles are expected to be ‘cold’, with a typical speed of $\sim 100 \text{ km s}^{-1}$ and they only interact weakly. For typical ranges of WIMP mass, 10 GeV to 10 TeV, the WIMP–nucleon interaction will be by elastic scattering. The signature of such an interaction would be nuclear recoil with energy in the range 1–100 keV. This energy is very low and requires specialized detection techniques. The cross sections are very small and so the event rates will be low. This presents an enormous experimental challenge to detect a signal above background. It is essential to reduce the rate of cosmic-ray interactions, which is

usually done by performing the experiments in deep underground caverns. The next background to consider is that from natural radioactivity. This can be reduced by very careful control of all the materials used in the detectors. Therefore, it is advantageous if the detector can have a powerful separation between the nuclear recoil signals and radioactive backgrounds. The detectors obviously have to be large to get a good sensitivity and there is no perfect solution to these challenges. There are currently many different approaches to these challenges. The following are some of the techniques used:

- **Scintillation detectors:** For example, the DAMA detector uses sodium iodide crystals as scintillators. Some separation between the signal from nuclear recoils and the backgrounds can be obtained from the measurement of the signal rise time. If there is a dark matter signal superimposed on a background, there should be an annual modulation because of the velocity of the Earth relative to the dark matter wind. The DAMA Collaboration have claimed to see such an effect, but the claim is controversial and has not been confirmed.
- **Semiconductor detectors:** These include germanium detectors. Again, the rise times can be used to provide some discrimination between nuclear recoils and radioactive backgrounds. Very low thresholds are required to see the small signals. If the detectors are cooled to very low temperatures, they can also be used as phonon detectors (see below) and the combination of the ionization and phonon signals provides powerful discrimination between nuclear recoils and backgrounds.
- **Phonon detectors:** Cryogenic detectors can be used to detect the nuclear recoil energy that is converted to phonons. One technique is to maintain a superconductor at the superconducting transition temperature so that a very small phonon signal will cause it to have a large and detectable change in resistance. If, in addition, the crystals are scintillating at cryogenic temperatures, the scintillation light can be read out using photomultipliers. This is very useful, because the correlation between the phonon and scintillation signal is very different for nuclear recoils and the backgrounds from β and γ radioactive sources. The CRESST experiment uses this technique with calcium tungstate crystals as the scintillator.
- **Noble liquids and gases:** Noble liquids such as xenon can be used to detect scintillation light. Dual noble liquid and gas detectors have also been used. A large electric field is used to drift the ionization electrons towards the gas in a wire detector. This allows the simultaneous measurement of the scintillation and ionization yield. Nuclear recoils from WIMPs are moving slowly and so tend to produce small direct ionization signals (S1) but they do produce scintillation light (S2) efficiently. The backgrounds from β and γ rays will also produce scintillation light, but they will also

produce larger ionization yields. Therefore, the ratio S2/S1 provides powerful discrimination between nuclear recoils (signal) and radioactive backgrounds. A diagram of the LUX detector [14] is shown in Fig. 13.29

Although there are several claims to have seen WIMP signals, they are all controversial and there are no signals confirmed by two experiments. Larger detectors are planned with target masses of the order of a ton, and these should be sensitive to WIMP signals over the range of WIMP masses and cross sections expected in common SUSY models that are compatible with the astrophysical data (see Section 13.7.1).

13.7.3 Dark matter annihilation

If dark matter accumulates under gravity in the centres of massive objects like the Sun, it should be possible to detect some of the products of the resulting annihilation reactions. One approach uses the muon neutrinos resulting from cascade decays. Neutrino ‘telescopes’ should then be able to detect the neutrinos and show that they come from the centre of the Sun. One such neutrino telescope is the IceCube [90], which has strings of photomultipliers buried in the Antarctic ice, giving it a volume of about 1 km^3 . Fast muons produce Čerenkov light, which is detected by the photomultipliers.³⁰ There is a background from downward-going muons, but upward-going muons can only come from neutrinos that have travelled through the Earth. Satellite experiments like FERMI, Pamela, and AMS are searching for the antiparticles that are expected from WIMP annihilation in the galactic halo. The measured positron-to-electron ratio is increasing with energy as would be expected in dark matter models, but the data are not yet able to exclude conventional sources such as pulsars.

13.8 Dark energy

Astronomers have been trying to measure the expansion of the Universe to determine if it is ‘closed’ (i.e. the expansion will be reversed at some future time) or ‘open’ (i.e. it will continue to expand indefinitely). This requires a determination of the rate of expansion with distance scale at the largest distances. The Universe is known to be expanding (called the ‘Hubble expansion’ after the astronomer who discovered this effect). The recession velocity of distant objects can be measured by looking at the redshifts of spectral lines compared with laboratory measurements of the rest values. The redshift is defined in terms of the shift in wavelength by $z = \Delta\lambda/\lambda$. The recession velocity β can be determined from the Doppler formula

$$1 + z = \sqrt{\frac{1 + \beta}{1 - \beta}} \quad (13.20)$$

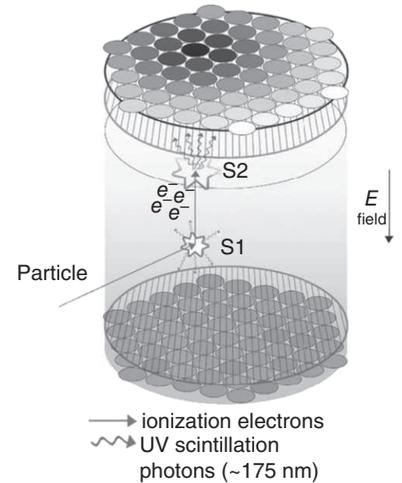


Fig. 13.29 LUX detector showing the arrays of photomultiplier tubes and the electrodes to create the required electric fields [14].

³⁰The cleanliness of the Antarctic ice results in a remarkably long attenuation length for the light of 55 m at a wavelength of 470 nm [89].

This picture is appropriate for nearby galaxies, but at larger distances Newtonian gravity is no longer correct. The Hubble law can be expressed as a linear relation between redshift and distance. The distances (called the luminosity distances D_L) are estimated by comparing the measured fluxes F and intrinsic luminosities L of specific objects for which L can be determined using the inverse square law

$$F = \frac{L}{4\pi D_L^2} \quad (13.21)$$

³¹This is referred to in astronomy as determining the intrinsic luminosity of ‘standard candles’. For example, the periods of Cepheid variable stars are correlated with their intrinsic luminosities. As the distances of nearby Cepheid variables can be determined geometrically by parallax measurements, this allows an absolute calibration.

³²Type Ia supernovae are believed to arise when a white dwarf star accretes enough mass from its main sequence (see Perkins in Further Reading if you are not familiar with this concept) companion that it exceeds the critical Chandrasekhar mass. The resulting nuclear fusion creates nickel and iron. The radioactive nickel atoms decay. It is assumed that more massive stars have to expand for longer before the opacity decreases enough for the photons to escape.

The main difficulty in this approach is the determination of L .³¹ For the largest distances the best method has used type Ia supernovae. Type Ia supernovae are bright enough to be detected at very large redshifts (and hence distances). By measuring the light output over a period of a few weeks, the width of the light curve can be measured. Empirically, for nearby supernovae for which the distances can be estimated with other techniques, there is a very good correlation between the width of the light curve and the intrinsic luminosity.³² This then allows type Ia supernovae to be used as standard candles. However, the further away and therefore older supernovae will have different chemical compositions compared with younger ones since they would have been formed from interstellar material that had not been through a cycle of nuclear fusion in stars. This effect might bias the calibration of the light curve to determine absolute luminosity. The resulting Hubble plot at large redshifts is shown in Fig. 13.30 and has clear deviations from linearity. The very surprising conclusion is that the expansion of the universe appears to be accelerating. More precise but less direct evidence for this acceleration can be determined from measurements of the anisotropy of the Cosmic Microwave Background (CMB) radiation.

13.8.1 Theoretical implications

One approach to understanding the accelerating expansion is to introduce a cosmological constant Λ into the Friedmann equation (see Perkins in Further Reading) for the expansion of the universe. The measured value corresponds to a value of approximately 5 GeV m^{-3} . In quantum field theory, we expect the vacuum fluctuations to contribute to the energy density. The integral for the energy density is divergent and if we assume there is a cut-off at the Planck scale, M_{Planck} , we would expect (see Exercise 13.10) that $\Lambda \sim 10^{121} \text{ GeV m}^{-3}$. New physics such as SUSY at the TeV scale would provide a much lower cut-off, but this would still disagree with the measured value by a factor of 10^{15} . One possible solution would be to assume that we need to modify Einstein’s theory of general relativity. Another option is that the cosmological constant does not arise from dark energy but from a new field called ‘quintessence’. A more radical alternative is based on the Anthropic Principle, according to which, in the context of inflationary models of the early universe, our universe might be just one universe in a larger multiverse and the

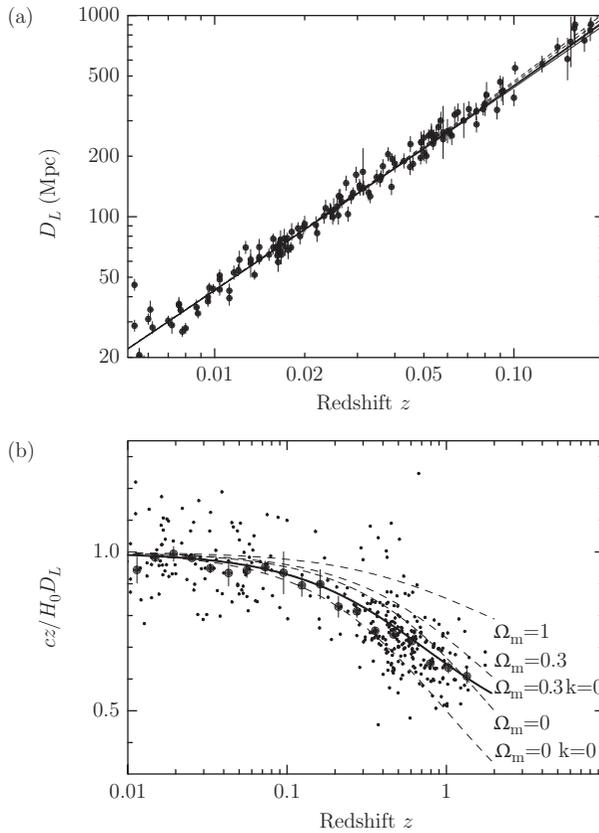


Fig. 13.30 (a) Hubble plot at moderate redshifts. (b) Hubble plot at large redshifts, with the linear trend seen at low z values divided out. The curves are fits to the Friedmann equation (see Perkins in Further Reading) with different parameters [115].

observed value of Λ might be a selection bias—only a universe with a sufficiently small value of Λ would be old enough to allow structures and life to evolve.

Chapter summary

- Many tests of the SM have already been performed at the LHC, giving confidence in the SM in this new regime of high CMS energy and low parton x .
- Many searches for BSM physics have so far proved negative, but a much larger search window is being opened up now that the LHC CMS energy has been raised to 13 TeV.
- A relatively low-energy e^+e^- collider would be able to make complementary, precision tests of the Higgs sector of the SM.
- There is astrophysical evidence for dark matter. This has led to a new area of research with the aim of direct or indirect detection of WIMPs.

Further reading

- Langacker, P. (2010). *The Standard Model and Beyond*. CRC Press. This has very thorough and advanced discussions of the SM and some BSM theories.
- Martin, S. P. (1997). *A Supersymmetry Primer*. arXiv:hep-ph/970935. This is a good introduction to the phenomenology and theory of SUSY.
- Aitchison, I. (2007). *Supersymmetry in Particle Physics: An Elementary Introduction*. Cambridge University Press. This gives a more advanced theoretical view of SUSY.
- Perkins, D. (2009). *Particle Astrophysics* (2nd edn). Oxford University Press. This is a very good introduction to many areas of physics at the interface between particle physics and astrophysics.

Exercises

- (13.1) By considering the combinatorics for one or two independent electron triggers to fire in $Z \rightarrow e^+e^-$ events, verify the formula for the trigger efficiency for the tag-and-probe analysis (eqn 13.1).
- (13.2) Consider the process $pp \rightarrow W \rightarrow l\nu$. If the W is produced with a mass M_W , determine an expression for the parton momentum fractions x_1 and x_2 of the protons in terms of the CMS energy \sqrt{s} . If the W is produced at rest, show that $x_1 = x_2$. Hence evaluate x_1 for this condition for the LHC operating at $\sqrt{s} = 7$ TeV.
- (13.3) What are the main SM backgrounds for a search involving final states with three charged leptons (as in the SUSY $\chi_2^0\chi_1^\pm$ production)? How could the SM background be reduced so as to enhance the sensitivity of a SUSY search? How could the SM background be estimated with minimal reliance on Monte Carlo calculations?
- (13.4) If a $\tilde{\gamma}$ has the same electromagnetic couplings as a γ , why does it have so much weaker interactions in matter?
- (13.5) Consider the elastic scattering of a (low-energy) WIMP with mass m_{WIMP} and momentum p with a nucleus of mass m_N . Assuming that the WIMP is non-relativistic (a very good approximation), show that the maximum kinetic energy of the recoil nucleus is given by
- $$T = \frac{2m_N p^2}{(m_N + m_{\text{WIMP}})^2}$$
- Discuss the implications for the choice of nuclear target for optimal sensitivity for a given WIMP mass. Evaluate the recoil energy for the case of WIMPs moving with a speed $v \sim 220 \text{ km s}^{-1}$, for a WIMP mass $m_{\text{WIMP}} = 1 \text{ GeV}$, assuming that the target nucleus is xenon (atomic mass 131). Repeat the calculation for a WIMP of mass $m_{\text{WIMP}} = 100 \text{ GeV}$. Discuss the implications for the direct detection of WIMPs and in particular of low-mass WIMPs.
- (13.6) Consider a dark matter search using a xenon (atomic mass 131 AMU) detector with a target mass of 10 tons. Assume that the local energy density of WIMPs is $\rho \sim 0.4 \text{ GeV cm}^{-3}$ and that the WIMP speed (relative to the nuclei) is $v = 220 \text{ km s}^{-1}$. If no events are observed in 1 year of operation, then, explaining any approximations you make, estimate a 90% confidence level upper limit to the WIMP–xenon cross section for a WIMP of mass 100 GeV. Why would a more accurate calculation result in less stringent limits.
- (13.7) Draw a Feynman diagram for the production of a W boson and a charm quark in pp interactions and hence explain how the measurement of this process can be used to determine the strange-quark parton distribution function at a value of $Q^2 = M_W^2$. Experimentally, it is found that the ratio of parton distribution functions (see eqn 9.56) $s(x, Q^2)/u_s(x)$ has a value of less than 1 at low values of Q^2 but is compatible with 1 at

$Q^2 = M_W^2$. If the charm quark produces a D meson, how might this process be tagged in an LHC experiment.

- (13.8) Draw a Feynman diagram for the production of prompt photons in pp interactions. Hence explain how the measured rate of prompt photon production at the LHC could be used to constrain the gluon parton density function.
- (13.9) Assuming that the angular distribution of jets ($dN/d\cos\theta^*$) in the parton-parton CMS is similar to that in Rutherford scattering, show that the distribution in χ (see eqn 13.17) will be approximately isotropic. Show that a uniform distribution in $\cos\theta^*$ will result in a peak in χ near $\chi = 1$. Explain from an experimental perspective the advantages to a search for new physics that uses the variable χ as opposed to the jet transverse momentum p_T .
- (13.10) The density of states for a quantum oscillator is $4\pi V k^2 dk / (2\pi)^3$, where V is the volume and k the momentum. Use this expression to evaluate the vacuum energy density, assuming that the divergent integral is cut off at some high energy scale E_{\max} . Estimate this value assuming that the cut-off is given by the Planck mass $M_{\text{Planck}} \sim 10^{19}$ GeV. Compare this value with the critical energy density $\rho_c \sim 5 \text{ GeV m}^{-3}$ and comment on the significance of this comparison. How would your conclusions change if we assumed that SUSY provided a cut-off at $E_{\max} \sim 1 \text{ TeV}$?

References

- [1] Aaltonen, T. et al. (2008). First run II measurement of the W boson mass at the Fermilab Tevatron. *Phys. Rev. D*, **77**, 112001.
- [2] Aaltonen, T. et al. (2012). Precision top-quark mass measurement at CDF. *Phys. Rev. Lett.*, **109**, 152003.
- [3] Aaltonen, T. et al. (2014). Precise measurement of the W -boson mass with the Collider Detector at Fermilab. *Phys. Rev. D*, **89**, 072003.
- [4] Aaron, F. D. et al. (2010). Combined measurement and QCD analysis of the inclusive $e^\pm p$ scattering cross sections at HERA. *JHEP*, **2010**, 109.
- [5] Abdallah, J. (2008). Measurement of the mass and width of the W boson in e^+e^- collisions at $\sqrt{s} = 161\text{--}209$ GeV. *EPJ C*, **55**, 1.
- [6] Abdesselam, A. et al. (2006). The barrel modules of the ATLAS semiconductor tracker. *Nucl. Instrum. Meth. A*, **568**, 642.
- [7] Abe, F. et al. (1995). Observation of top quark production in $\bar{p}p$ collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, **74**, 2626.
- [8] Abe, K. et al. (2014). Observation of electron neutrino appearance in a muon neutrino beam. *Phys. Rev. Lett.*, **112**, 061802.
- [9] Abrams, G. S. et al. (1975). Decay of $\Psi(3684)$ into $\Psi(3095)$. *Phys. Rev. Lett.*, **34**, 1181.
- [10] Adamson, P. et al. (2011). Measurement of the neutrino mass splitting and flavor mixing by MINOS. *Phys. Rev. Lett.*, **106**, 181801.
- [11] Adler, S. L. (1966). Sum rules giving tests of local current commutation relations in high-energy neutrino reactions. *Phys. Rev.*, **143**, 1144.
- [12] Aharmim, B. et al. (2005). Electron energy spectra, fluxes, and day–night asymmetries of ^8B solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory. *Phys. Rev. C*, **72**, 055502.
- [13] Aitchison, I. J. R. and Hey, A. J. G. (2013). *Gauge Theories in Particle Physics*, Volumes 1 and 2 (4th edn). CRC Press.
- [14] Akerib, D. S. et al. (2013). The Large Underground Xenon (LUX) experiment. *Nucl. Instrum. Meth. A*, **704**, 111.
- [15] Albajar, C. et al. (1989). Studies of intermediate vector boson production and decay in UA1 at the CERN proton–antiproton collider. *Z. Phys. C*, **44**, 15.
- [16] ALEPH, DELPHI, L3, and OPAL Collaborations (2001). Summer 01 LEP 4f averages: WW cross-section vs sqrt s for different W couplings. <http://lepewwg.web.cern.ch/LEPEWWG/lepww/4f/Summer01/>.

- [17] ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups (2006). Precision electroweak measurements on the Z resonance. *Phys. Rept.*, **427**, 257.
- [18] Amaudruz, P. et al. (1991). Gottfried sum from the ratio F_2^n/F_2^p . *Phys. Rev. Lett.*, **66**, 2712.
- [19] ATLAS, CDF, CMS, D0 Collaborations (2014). First combination of Tevatron and LHC measurements of the top-quark mass. arXiv:1403.4427 [hep-ex].
- [20] ATLAS Collaboration (2008). The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, **3**, S08003.
- [21] ATLAS Collaboration (2011). An extrapolation to a larger fiducial volume of the measurement of the $W \rightarrow l\nu$ charge asymmetry in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. ATLAS-CONF-2011-129. <http://cds.cern.ch/record/1383832>.
- [22] ATLAS Collaboration (2012). A measurement of the inclusive W^\pm and Z/γ^* cross sections in the e and μ decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. D*, **85**, 072004.
- [23] ATLAS Collaboration (2012). A measurement of the inclusive W^\pm and Z/γ^* cross sections in the e and μ decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector (auxiliary material). <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2011-06/>.
- [24] ATLAS Collaboration (2012). Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev. D*, **86**, 014022.
- [25] ATLAS Collaboration (2012). Measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using final states with two high- p_T leptons. *JHEP*, **05**, 059.
- [26] ATLAS Collaboration (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, **716**, 1.
- [27] ATLAS Collaboration (2012). Search for high-mass resonances decaying to dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, **11**, 138.
- [28] ATLAS Collaboration (2013). Evidence for Higgs boson decays to the $\tau^+\tau^-$ final state with the ATLAS detector. ATLAS-CONF-2013-108. <http://cds.cern.ch/record/1632191>.
- [29] ATLAS Collaboration (2013). Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, **726**, 120.
- [30] ATLAS Collaboration (2013). Measurements of Higgs production and couplings using diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B*, **726**, 88.
- [31] ATLAS Collaboration (2013). Search for new phenomena in final states with large jet multiplicities and missing transverse momentum with proton–proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS experiment. *JHEP*, **10**, 130.
- [32] ATLAS Collaboration (2013). Search for new physics in dijet mass and angular distributions in pp collisions at $\sqrt{s} = 7$ TeV measured with the ATLAS detector. *JHEP*, **01**, 029.
- [33] ATLAS Collaboration (2014). Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b -tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *EPJ C*, **74**, 3109.

- [34] ATLAS Collaboration (2014). Search for direct pair production of the top squark in all-hadronic final states in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, **09**, 015.
- [35] ATLAS Collaboration (2014). Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, **05**, 071.
- [36] ATLAS Collaboration (2014). Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using $\sqrt{s} = 8$ TeV proton–proton collision data. *JHEP*, **09**, 176.
- [37] ATLAS Collaboration (2015). Probing the measurement of jet energies with the ATLAS detector using photon + jet events in proton–proton collisions at $\sqrt{s} = 7$ TeV.
- [38] Aubert, B. et al. (1991). Performance of a liquid argon electromagnetic calorimeter with an accordion geometry. *Nucl. Instrum. Meth. A*, **309**, 438.
- [39] Aubert, J. J. et al. (1974). Experimental observation of a heavy particle J . *Phys. Rev. Lett.*, **33**, 1404.
- [40] Augustin, J. et al. (1974). Discovery of a narrow resonance in e^+e^- annihilation. *Phys. Rev. Lett.*, **33**, 1406.
- [41] BaBar Collaboration (2009). Dalitz plot analysis of $D_s^+\pi^+\pi^-\pi^+$. *Phys. Rev. D*, **79**, 032003.
- [42] Bahcall, J. N., Serenelli, A. M., and Basu, S. (2005). New solar opacities, abundances, helioseismology, and neutrino fluxes. *Astrophys. J.*, **621**, L85.
- [43] Barate, R. et al. (2000). Measurement of the Z resonance parameters at LEP. *EPJ C*, **14**, 1.
- [44] Barnes, V. E. et al. (1964). Confirmation of the existence of the Ω^- hyperon. *Phys. Lett.*, **12**, 134.
- [45] Bartel, W. et al. (1980). Observation of planar three-jet events in e^+e^- annihilation and evidence for gluon bremsstrahlung. *Phys. Lett. B*, **91**, 142.
- [46] Barut, A. O. and R., Rączka (1986). *Theory of Group Representations and Applications* (2nd revised edn). World Scientific.
- [47] Begeman, K. G. (1991). Extended rotation curves of spiral galaxies—dark haloes and modified dynamics. *MNRAS*, **249**, 523.
- [48] Belle Collaboration (2012). Precise measurement of the CP violation parameter $\sin 2\phi_1$ in $B^0 \rightarrow c\bar{c}K^0$ decays. *Phys. Rev. Lett.*, **108**, 171802.
- [49] Bennett, G. W. et al. (2006). Final report of the E821 muon anomalous magnetic moment measurement at BNL. *Phys. Rev. D*, **73**, 072003.
- [50] Berestetskii, W. B., Lifshitz, E. M., and Pitaevskii, L. P. (1971). *Relativistic Quantum Theory, Part I*. Pergamon Press.
- [51] Bernstein, A. et al. (1993). Beam tests of the ZEUS barrel calorimeter. *Nucl. Instrum. Meth. A*, **336**, 23.
- [52] Bjorken, J. and Drell, S. D. (1964). *Relativistic Quantum Mechanics*. McGraw-Hill.
- [53] Bohr, A. and Mottelson, B. (1998). *Nuclear Structure I*. World Scientific (First published in 1969 by W. A. Benjamin).
- [54] Brandelik, R. et al. (1980). Evidence for a spin-1 gluon in three-jet events. *Phys. Lett. B*, **97**, 453.

- [55] Brandelik, R. et al. (1982). Charge asymmetry and weak interaction effects in $e^+e^- \rightarrow \mu^+\mu^-$ and $e^+e^- \rightarrow \tau^+\tau^-$. *Phys. Lett. B*, **110**, 173.
- [56] Brown, C. N. et al. (1989). Dimuon production in 800-GeV proton–nucleus collisions. *Phys. Rev. Lett.*, **63**, 2637.
- [57] Burcham, W. E. and Jobes, M. (1994). *Nuclear and Particle Physics*. Pearson.
- [58] Cacciari, M. et al. (2008). The anti- k_T jet clustering algorithm. *JHEP*, **04**, 063.
- [59] Callan, C. G. and Gross, D. (1969). High-energy electroproduction and constitution of electric current. *Phys. Rev. Lett.*, **22**, 156.
- [60] Chambers, R. G. (1960). Shift of an electron interference pattern by enclosed magnetic flux. *Phys. Rev. Lett.*, **5**, 3.
- [61] Christenson, J. H. et al. (1964). Evidence for the 2π decay of the K_2^0 meson. *Phys. Rev. Lett.*, **13**, 138.
- [62] CMS Collaboration (2008). The CMS Experiment at the CERN LHC. *JINST*, **3**, S08004.
- [63] CMS Collaboration (2011). Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, **01**, 080.
- [64] CMS Collaboration (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, **716**, 30.
- [65] CMS Collaboration (2013). Study of the mass and spin–parity of the Higgs boson candidate via its decays to Z boson pairs. *Phys. Rev. Lett.*, **110**, 081803.
- [66] CMS Collaboration (2014). Search for the Standard-Model Higgs boson decaying to a pair of τ leptons. *JHEP*, **05**, 104.
- [67] CMS Collaboration (2015). Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the Standard Model predictions using proton collisions at 7 and 8 TeV. *EPJ C*, **75**, 212.
- [68] Condon, E. U. and Shortley, G. H. (1935). *The Theory of Atomic Spectra*. Cambridge University Press.
- [69] Danby, G. et al. (1962). Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, **9**, 36.
- [70] de Boer, W. and Sander, C. (2003). Global electroweak fits and gauge coupling unification. arXiv: hep-ph/0307049.
- [71] de Groot, J. G. H. et al. (1979). Inclusive interactions of high-energy neutrinos and antineutrinos in iron. *Z. Phys. C*, **1**, 143.
- [72] DELPHI Collaboration (1997). Measurement of $B_d^0-\bar{B}_d^0$ oscillations. *Z. Phys. C*, **76**, 579.
- [73] Evans, L. and Bryant, P. (2008). The LHC machine. *JINST*, **3**, S08001.
- [74] Feynman, R. P. (1987). The reason for antiparticles. In Feynman, R. P. and Weinberg, S. *Elementary Particles and the Laws of Physics: The 1986 Dirac Memorial Lectures*. Cambridge University Press.
- [75] Feynman, R. P., Leighton, R. B., and Sands, M. (1965). *Feynman Lectures on Physics*. Volume II. Addison-Wesley.
- [76] Fong, C. and Kittel, C (1967). Induced charge on capacitor plates. *Am. J. Phys.*, **35**, 1091.

- [77] Gando, A. et al. (2011). Constraints on θ_{13} from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND. *Phys. Rev. D*, **83**, 052002.
- [78] Glashow, S., Iliopoulos, J., and Maiani, L. (1970). Weak interactions with lepton–hadron symmetry. *Phys. Rev. D*, **2**, 1285.
- [79] Goldhaber, M., Grodzins, L., and Sunyar, W. (1958). Helicity of neutrinos. *Phys. Rev.*, **109**, 1015.
- [80] Griffiths, D. (2008). *Introduction to Elementary Particles* (2nd revised edn). Wiley-VCH.
- [81] Gross, D. J. and Llewellyn Smith, C. H. (1969). High-energy neutrino–nucleon scattering, current algebra and partons. *Nucl. Phys. B*, **14**, 337.
- [82] H1 and ZEUS Collaborations. Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data. arXiv:1506.06042 [hep-ex].
- [83] Haber, H. E. (2014). Supersymmetry, Part 1 (Theory). In Review of Particle Physics (Particle Data Group). *Chin. Phys. C*, **38**, 1554. <http://hepdata.cedar.ac.uk/lbl/2014/reviews/rpp2014-rev-susy-1-theory.pdf>.
- [84] Halzen, F. and Martin, A. D. (1984). *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley.
- [85] Hanneke, D., Fogwell Hoogerheide, S., and Gabrielse, G. (2011). Cavity control of a single-electron quantum cyclotron: measuring the electron magnetic moment. *Phys. Rev. A*, **83**, 052122.
- [86] Herb, S. W. et al. (1977). Observation of a dimuon resonance at 9.5 GeV in 400 GeV proton–nucleus collisions. *Phys. Rev. Lett.*, **39**, 252.
- [87] Hofstadter, R. (1956). Electron scattering and nuclear structure. *Rev. Mod. Phys.*, **28**, 214.
- [88] Holder, M. et al. (1978). Performance of a magnetized total absorption calorimeter between 15 GeV and 140 GeV. *Nucl. Instrum. Meth. A*, **151**, 69.
- [89] IceCube Collaboration (2001). IceCube Preliminary Design Document. <http://icecube.wisc.edu/icecube/static/reports/IceCubeDesignDoc.pdf>.
- [90] IceCube South Pole Neutrino Observatory. <http://icecube.wisc.edu/>.
- [91] Kluth, S. (2006). Tests of quantum chromo dynamics at e^+e^- colliders. *Rep. Prog. Phys.*, **69**, 1771.
- [92] Kodama, K. et al. (2001). Observation of tau neutrino interactions. *Phys. Lett. B*, **504**, 218.
- [93] KTeV Collaboration (2011). Precise measurements of direct CP violation, CPT symmetry, and other parameters in the neutral kaon system. *Phys. Rev. D*, **83**, 092001.
- [94] Landau, L. D. and Lifshitz, E. M. (1977). *Quantum Mechanics (Non-Relativistic Theory)* (3rd edn). Pergamon Press.
- [95] Landau, L. D. and Yang, C. N. (1950). Selection rules for the dematerialization of a particle into two photons. *Phys. Rev.*, **77**, 242.
- [96] Landau, R. H. (1996). *Quantum Mechanics II: A Second Course in Quantum Theory* (2nd edn). Wiley.
- [97] LEP Electroweak Working Group (2012). LEP/TEV EW WG Plots for Winter 2012: w12_mt_mw_contours.pdf. <http://lepewwg.web.cern.ch/LEPEWWG/plots/winter2012/>.

- [98] Leung, W. C. et al. (1993). A measurement of the Gross–Llewellyn Smith sum rule from the CCFR xF_3 structure function. *Phys. Lett. B*, **317**, 655.
- [99] LHCb Collaboration (2008). LHCb experiment at the CERN LHC. *JINST*, **3**, S08005.
- [100] LHCb Collaboration (2011). Measurement of the inclusive ϕ cross-section in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, **703**, 267.
- [101] LHCb Collaboration (2013). First observation of CP violation in the decays of B_s^0 mesons. *Phys. Rev. Lett.*, **110**, 221601.
- [102] LHCb Collaboration (2013). Measurement of the time-dependent CP asymmetry in $B^0 \rightarrow J/\Psi K_s^0$ decays. *Phys. Lett. B*, **721**, 24.
- [103] LHCb Collaboration (2013). Observation of D^0 – \bar{D}^0 oscillations. *Phys. Rev. Lett.*, **110**, 101802.
- [104] LHCb Collaboration (2013). Precision measurement of the B_s^0 – \bar{B}_s^0 oscillation frequency with the decays $B_s^0 \rightarrow D_s^- \pi^+$. *New J. Phys.*, **15**, 053021.
- [105] LHCb Collaboration (2014). Observation of $B_s^0 \rightarrow K^* K^\pm$ and evidence for $B_s^0 \rightarrow K^{*-} \pi^+$ decays. *New J. Phys.*, **16**, 123001.
- [106] LHCb Collaboration (2015). Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \rightarrow J/\psi$ decays. *Phys. Rev. Lett.*, **115**, 072001.
- [107] Maggiore, M. (2005). *A Modern Introduction to Quantum Field Theory*. Oxford University Press.
- [108] Mikheyev, S. and Smirnov, A. (1986). Resonant amplification of ν -oscillations in matter and solar-neutrino spectroscopy. *Nuovo Cim. C*, **9**, 17.
- [109] MINOS Collaboration (2014). The MINOS Experiment and NuMI Beamline. <http://www.numi.fnal.gov/>.
- [110] Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973). *Gravitation*. W. H. Freeman.
- [111] Moriyasu, K. (1983). *An Elementary Primer for Gauge Theory*. World Scientific.
- [112] NA31 Collaboration (1990). A measurement of the phases of the CP violating amplitudes in $K^0 \rightarrow 2\pi$ decays and a test of CPT invariance. *Phys. Lett. B*, **237**, 303.
- [113] NA48 Collaboration (2007). The beam and detector for the NA48 neutral kaon CP violations experiment at CERN. *Nucl. Instrum. Meth. A*, **574**, 433.
- [114] Particle Data Group (2010). Review of Particle Physics. *J. Phys. G*, **37**, 075021.
- [115] Particle Data Group (2014). Review of Particle Physics. *Chin. Phys. C*, **38**, 090001.
- [116] Patterson, J. R. et al. (1990). A determination of $\text{Re}(\epsilon'/\epsilon)$ by the simultaneous detection of the four $K_{L,S} \rightarrow \pi\pi$ decay modes. *Phys. Rev. Lett.*, **64**, 1491.
- [117] Perkins, D. H. (1987). *Introduction to High Energy Physics* (3rd edn). Addison-Wesley.
- [118] Perkins, D. H. (2000). *Introduction to High Energy Physics* (4th edn). Cambridge University Press.
- [119] Plehn, T. http://www.thphys.uni-heidelberg.de/~plehn/includes/prospino/prospino_lhc8.eps.
- [120] Ramsey, N. F. (1980). The method of successive oscillatory fields. *Physics Today*, **33**, 25.
- [121] Ramsey, N. F. (1990). Electric dipole moment of the neutron. *Annu. Rev. Nucl. Particle Sci.*, **40**, 1.

- [122] Rauch, H. et al. (1975). Verification of coherent spinor rotation of fermions. *Phys. Lett. A*, **54**, 523.
- [123] Rochester, G. D. and Butler, C. C. (1947). Evidence for the existence of new unstable elementary particles. *Nature*, **160**, 855.
- [124] Sakharov, A. D. (1991). Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe. *Sov. Phys. Usp.*, **34**, 392.
- [125] Sauli, F. (1977). Principles of operation of multiwire proportional and drift chambers. CERN Report 77-09. http://lhcb-muon.web.cern.ch/lhcb-muon/documents/Sauli_77-09.pdf.
- [126] Schiff, L. I. (1968). *Quantum Mechanics* (3rd edn). McGraw-Hill.
- [127] Schutz, B. F. (1985). *A First Course in General Relativity*. Cambridge University Press.
- [128] Steane, A. M. (2012). *Relativity Made Relatively Easy*. Oxford University Press.
- [129] Super-Kamiokande Collaboration. Images of Super-Kamiokande events from tscan. <http://www.ps.uci.edu/~tomba/sk/tscan/pictures.html>.
- [130] 't Hooft, G. and Veltman, M. (1972). Regularization and renormalization of gauge fields. *Nucl. Phys. B*, **44**, 189.
- [131] Tevatron Electroweak Working Group. Combination of CDF and D0 results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron. arXiv:1305.3929 [hep-ex].
- [132] Van der Meer, S. (1985). Stochastic cooling and the accumulation of antiprotons. *Science*, **230**, 900.
- [133] Vilain, P. et al. (1994). Measurement of electroweak parameters from the scattering of muon-neutrinos on electrons. *Phys. Lett. B*, **335**, 246.
- [134] Winter, K. (2000). *Neutrino Physics* (2nd edn). Cambridge University Press.
- [135] Wolfenstein, L. (1978). Neutrino oscillations in matter. *Phys. Rev. D*, **17**, 2369.
- [136] Wu, S. L. (1984). e^+e^- physics at PETRA—the first five years. *Phys. Rept.*, **107**, 59.

Index

A

- Accelerator physics
 - Q value of cavity, 51
 - bucket, 49
 - bunches, 49
 - cavity breakdown limit, 53
 - chain of accelerators, 59
 - energy gain, 50
 - FODO structure, 56
 - klystron, 54
 - phase stability, 50
 - strong focusing, 56
 - superconducting cavities, 53
 - synchrotron oscillations, 50
- Accelerators
 - brief history, 8
 - colliding-beam experiments, 62
- Angular momentum, 19
 - addition, 21
 - Clebsch–Gordan coefficients, 21
 - raising and lowering operators, 20

B

- Beam optics, 54
 - beam blow-up, 58
 - betatron oscillations, 58
 - dipole bending strength, 55
 - emittance, 58
 - Hill's equations, 57
 - Liouville's theorem, 58
 - quadrupole magnet focal length, 56
 - quadrupole magnets, 55
 - reference trajectory, 54
 - transverse beam phase space, 59
- Beyond the Standard Model (BSM)
 - LHC trigger
 - missing energy, 359
 - overview, 358
 - SM cross sections for Tevatron and LHC, 356
 - supersymmetry (SUSY), 368
- Beyond the Standard Model (BSM) physics, 367
 - dark energy, 385
 - theoretical implications, 386
 - dark matter, 381
 - astrophysical evidence, 382
 - detection, 383
 - Experimental searches at the LHC
 - SUSY, 373

- experimental searches at the LHC, 373
- hierarchy problem, 367
- linear collider and Higgs physics, 379
- other BSM theories, 372
- SUSY
 - R -parity, 371
 - coupling constant evolution, 370
 - lightest supersymmetric particle (LSP), 370
- Breit–Wigner
 - branching ratios, 42
 - formula, 40
 - maximum cross section, 42
 - spin factor, 41

C

- Charged-current weak interaction, 181
 - left-handed particles, right-handed antiparticles, 186
- CKM matrix, 190
 - measurement of matrix elements, 207
- CMS \leftrightarrow laboratory-frame energies and angles, 241
- Colliders
 - $\bar{p}p$, 65
 - HERA ep , 63
 - LEP e^+e^- , 63
 - LHC pp , 63
- Colour charge and gluons, 7
- CP violation
 - B^0 system with mixing
 - formalism, 296
 - time evolution, 297
 - K_L semileptonic charge asymmetry, 291
 - $\bar{\epsilon}$ parameter, 291
 - mathematical formalism, 291
- B-factories, 299
 - $\Upsilon(4S)$, 299
 - $\Upsilon(4S)$, A_{CP} , 300
 - beyond CKM description
 - 'penguin' diagram, 296
- CKM matrix, 294
 - small for K^0 , D^0 , large for B^0 , 295
 - unitarity constraints and triangles, 295
- CKM unitarity triangle
 - Jarlskog invariant, 295
 - constraining the CKM phase δ , 297

- discovery, 289
 - in K^0 decay
 - $R \neq 1$ from combined measurements, 294
 - K^0 decay
 - double ratio R , 292
 - parameters η_{00} , η_{+-} , 292
 - LHCb detector, 301
 - LHCb measurements, 302
 - NA48 experiment, 292
 - time-dependent asymmetries, 297
 - expression for $A_{CP}(t)$, 299
 - Cross section, 39
 - dLips for n particles, 39
 - particle flux factor, 40
- ## D
- $\Delta \rightarrow \pi$ -nucleon branching ratios, 121
 - Data analysis
 - W^+W^- at LEP2
 - W width measurement, 223
 - calibration, 213
 - Monte Carlo simulation, 213
 - W^+W^- at LEP2, 222
 - W mass and width, 223
 - choice of final state, 222
 - combinatorics, 222
 - Z^0 at LEP1, 213
 - Decay rate, 34
 - three-body decay, 36
 - Dalitz plot, 38
 - two-body decay, 34
 - two-body kinematics, 36
 - Decay rate and lifetime, 31
 - Deep inelastic scattering (DIS), 235
 - $\nu_e e$ elastic scattering, 243
 - $\bar{\nu}_e e$ elastic scattering
 - differential cross-section, 244
 - scattering amplitude, 243
 - charged-lepton probes, 251
 - $e\mu$ cross section, 252, 253
 - $e\mu$ elastic scattering, 252
 - Bjorken scaling, 255
 - Callan–Gross relation, 254
 - cross section in terms of PDFs, 254
 - cross section in terms of structure functions, 254
 - electron–nucleon scattering, 254
 - electron–quark scattering, 254
 - Gottfried sum rule, 258

- Deep inelastic scattering (DIS)
 - (*continued*)
 - electroweak unification at HERA:
 - neutral- and charged-current DIS at high Q^2 , 258
 - neutrino probes, 242
 - V–A interaction, 242, 243
 - neutrino–nucleon cross sections, 246
 - neutrino–quark scattering, 245
 - kinematics, 245
 - structure functions
 - $\frac{5}{18}$ ratio from quark charges, 257
 - Detectors, 71
 - RICH detector, 105
 - charged-particle momentum determination, 97
 - charged-particle trackers, 86
 - drift chambers, 90
 - gas gain, 87
 - proportional wire chambers, 89
 - collider overview, 72
 - complete systems
 - ATLAS, 109
 - CMS, 110
 - electromagnetic calorimeter, 98
 - energy resolution, 102
 - homogeneous, 99
 - sandwich, 100
 - for particle identification, 105
 - Čerenkov radiation, 105
 - RICH detector, 105
 - transition radiation, 105
 - using ionization, 106
 - hadronic calorimeter, 103
 - e/h compensation, 103
 - energy resolution, 105
 - magnetic fields
 - charged-particle tracking, 106
 - muon spectrometer, 107
 - neutrinos, 111
 - particle jets, 98
 - photon detection, 85
 - avalanche photodiode, 86
 - photomultiplier, 85
 - signal generation, 81
 - induced current, 83
 - moving charges, 81
 - Ramo’s theorem, 84
 - scintillators, 84
 - silicon detectors, 92
 - bias voltage, 93
 - design parameters, 92
 - drift velocity, 93
 - pixel systems, 96
 - radiation damage, 94
 - strip systems, 95
 - trigger, 108
 - levels, 108
 - LHC trigger, 109
 - wire chambers
 - drift velocity, 91
 - signal and readout, 90
 - signal pulse shape, 92
 - signal readout circuit, 92
 - Diagrams, 7
 - Feynman diagrams, 7
 - quark flow diagrams, 8
 - Dirac equation, 155
 - Ψ 4-component Dirac spinor, 156
 - α , β matrices
 - Weyl representation, 157
 - γ matrices
 - Weyl representation, 156
 - Foldy–Wouthuysen (FW) representation, 158
 - decoupled equations for $E > 0$, $E < 0$, 158
 - bilinear covariant, 157
 - charge conjugation
 - particle \leftrightarrow antiparticle symmetry, 168
 - covariant form, 156
 - γ matrices, 156
 - adjoint equation, 156
 - crossing symmetry
 - pair creation, annihilation, 165
 - Electromagnetic interactions, 169
 - minimal coupling, 169
 - free-particle solutions, 159
 - $\Psi^+(x)$, for $E > 0$, 160
 - $\Psi^-(x)$, for $E < 0$, 160
 - frame with $p > 0$, 160
 - chirality operators P_L, P_R , 161
 - helicity eigenvalues, 161
 - helicity operator, 161
 - rest frame, 159
 - Hamiltonian, 157
 - helicity \neq chirality, 163
 - $\pi \rightarrow \mu\nu$ decay, 163
 - high energies
 - helicity \equiv chirality, 164
 - current–current interaction, 164
 - helicity conservation, 165
 - insight into origin, 146
 - Lorentz-covariant one-particle states, 147
 - Majorana particles, 159
 - non-relativistic limit, 170
 - equation for Ψ_U , 170
 - magnetic moment of electron, 171
 - Pauli Hamiltonian, 171
 - upper and lower components of Ψ , 170
 - probability current and density, 157
 - space inversion
 - operator, 166
 - parity for $E > 0$, $E < 0$ states, 167
 - standard representation, 158
 - non-relativistic limit, 158
 - time inversion, 167
 - anti-unitary operator, 168
 - Kramers degeneracy, 168
 - unitary transformation between representations, 158
- Discrete symmetries, 15
 - J^{PC} of hadrons, 18
 - charge conjugation, 16
 - parity, 15
 - time reversal, 17
- dLips (Lorentz-invariant phase space), 33
- E**
 - Electroweak unification, 193
 - W^- - and Z -boson masses, 200
 - W^0, B^0 , 196
 - electroweak properties of fermions, 199
 - How good is it?, 201
 - left-handed particles, 194
 - neutral-current couplings c_V, c_A , 198
 - neutral-current couplings g_L, g_R , 198
 - neutral-current interaction, 198
 - unification condition, 198
 - weak isospin, 195
 - weak mixing angle (Weinberg angle), 196
 - weak neutral currents, 199
- F**
 - Fermi constant G_F , 185
 - Feynman rules, 183
 - Feynman–Stueckelberg, $E < 0$ states, 151
 - Fixed-target experiments, 62
 - Flavour oscillations
 - B_s , 287
 - detector requirements, 287
 - $B^0-\bar{B}^0$, 285
 - $D^0-\bar{D}^0$, 284
 - $K^0-\bar{K}^0$, 283
 - time dependence different Δm , 283
 - Forces, 6
- G**
 - Gauge symmetry, 172
 - Aharonov-Bohm effect, 174
 - classical electromagnetism, 174
 - gauge transformation of potentials, 174
 - Covariant derivative, 172
 - interactions from gauge symmetry, 175
 - covariant derivative, 175
 - Dirac equation and
 - electromagnetic field, 175
 - general formalism, 177
 - GIM mechanism, 189

- Group theory, 43
 - $U(n)$ and $SU(n)$, 44
 - in particle physics, 4
 - Lie groups, 44
 - Representation theory
 - combining states, 46
 - $SU(2)$, 45
 - Pauli matrices, 45
- H**
- Hadron colliders
 - W and Z discovery, 224
 - E_T^{miss} , 224
 - top-quark mass (CDF and D0), 231
 - top-quark physics, 229
 - discovery, 229
- HERA ep collider, 63
- Higgs discovery, 342
 - WW^* channel, 346
 - ZZ^* channel, 345
 - $\gamma\gamma$ channel, 344
 - π^0 decay background, 344
 - decay modes and branching ratios, 343
 - expected mass, 342
 - fermion decay modes, 349
 - outlook for new measurements, 353
 - production mechanism, 343
 - statistical significance, 347
- Higgs mechanism, 333
 - classical fields, 334
 - criteria for the Higgs field, 334
 - key ideas, 336
 - Lagrangians, 339
 - final form with massive fields, 341
 - local gauge invariance, 334
 - mathematical approach, 339
 - spontaneous symmetry breaking, 335
 - hidden symmetry, 336
 - Mexican hat potential, 336
- I**
- Ionization energy loss, 74
 - Bethe formula, 75
- Isoscalar nuclear target, 249
- J**
- Jet-finding algorithms, 360
 - infrared safety, 360
 - jet energy calibration, 361
- K**
- K^0, \bar{K}^0
 - K_1, K_2 , 278
 - lifetimes and decay modes, 280
 - time dependence of beams, 280
 - K_1, K_2 CP eigenstates, 279
 - $\pi^+\pi^-, \pi^+\pi^-\pi^0$ decay modes, 278
 - CP values, 279
- mixing
 - box diagrams, 279
 - regeneration, 288
 - semileptonic decays, 278
 - strangeness indicator, 278
- K_S, K_L physical states, short and long lifetimes, 289
- KamLAND $\bar{\nu}_e$ experiment, 322
- Klein–Gordon equation, 149
 - $E < 0$ solutions, 149
 - $E > 0$ solutions, 149
 - inclusion of a potential, 152
 - $V > 0$ barrier in one dimension, 152
 - probability current and density, 150
 - $E < 0$, 150
 - $E > 0$, 150
 - normalization, 150
- Klystron, 54
- L**
- Lagrangians, 337
 - classical mechanics, 337
 - conserved quantities, 338
 - effective potential, 338
 - Euler–Lagrange equations, 338
 - independent variables, 337
 - local gauge invariance, 341
 - choice of gauge, 341
 - mass generation, 340
- LEP e^+e^- collider, 63
- Lepton number, 183
- LHC
 - search for physics Beyond the Standard Model (BSM), 356
- LHC pp collider, 63
- LHC dipole magnets, 59
- LHC SM measurements, 359
 - $W \rightarrow \mu\nu_\mu E_T^{\text{miss}}$ distribution from CMS, 362
 - W charge asymmetry from $W \rightarrow l\nu$ decays, 363
 - Z peak, e^+e^- invariant mass spectrum, 362
 - cross-section formula, 362
 - jet cross sections, 359
 - top-quark production, 365
- LHC top-quark production
 - event selection, 365
- LHCb detector, 301
- Linear accelerator, 53
 - disc-loaded structure, 54
 - travelling-wave structure, 53
- Lorentz invariance, 29
 - invariant variables, 30
 - rapidity, 30
- Lorentz transformation
 - spin-matrix form
 - boost and rotation, 142
 - boost only, 142
- Luminosity, 63
 - accelerator parameters, 65
 - Gaussian beams, 64
 - optimization, 65
- M**
- Majorana particles, 159
- MINOS experiment
 - ν_μ flux
 - far detection by charged-current interaction, 315
 - near detector, 315
- MSW matter effect
 - neutrinos in matter, 320
- N**
- ν_μ beam
 - production and focusing, 314
- Natural units, 2
- Negative-energy states in relativistic quantum mechanics, 151
- Neutral meson systems
 - flavour oscillations, 282
 - quantum mechanics, mass matrix, 281
 - summary table ($\Delta m, \tau$), 278
- Neutral-current weak interaction, 181
- Neutrino states
 - PMNS mixing matrix, 307
 - weak and mass, 307
- Neutrino structure functions F_2, F_3
 - isoscalar nuclear-target PDFs, 249
 - neutron-target PDFs, 249
 - proton-target PDFs, 249
- Neutrinos
 - $\nu_e \neq \nu_\mu$ and lepton number, 206
 - DONUT ν_τ direct evidence, 206
 - atmospheric, 311
 - ν_μ -to- ν_e ratio, 312
 - double ratio
 - $R' = R_{\text{experiment}}/R_{\text{predicted}}$, 312
 - Ratio $R = \text{Flux}(\nu_\mu)/\text{Flux}(\nu_e)$, 312
 - Super-Kamiokande Čerenkov detector, 312
 - zenith-angle distributions, 314
 - detection
 - $(\Delta m)^2$ sensitivity summary table, 311
 - the experimental challenge, 310
 - first evidence, 205
 - helicity measurement (Goldhaber), 206
 - mass upper limits, 306
 - double β decay, 307
 - tritium β decay, 306
 - matter–antimatter asymmetry, 329
 - Sakharov conditions, 329
 - MINOS ν oscillation experiment, 315

Neutrinos (*continued*)

- Solar Neutrino Oscillation (SNO) experiment, 317
 - solar neutrinos, 315
 - three (or more)-flavour oscillations, 323
 - CP and T violation, 325
 - $P(\nu_\alpha \rightarrow \nu_\beta)$ at distance L , 324
 - $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$ at distance L , 325
 - time dependence of ν mass eigenstates, 324
 - three-flavour oscillations
 - CP violation, 327
 - Δm_{ij}^2 ordering, 326
 - mixing angle θ_{13} , 327
 - mixing matrix, 326
 - PMNS mixing matrix, 326
 - two-flavour oscillations, 308
 - $P(\alpha \rightarrow \beta)$ at distance L , 309
 - factor ‘1.27’ units, 310
 - one mixing angle, 308
 - phase angle ϕ , 309
 - probabilities for appearance and survival, 309
 - time evolution, 308
 - V–A parity-violating ^{60}Co β decay, 206
- NuMI beam
- high-energy ν_μ , 314

O

- Ω^- , $S = -3$, spin- $\frac{3}{2}$ discovery, 127

P

- $\bar{p}p$ colliders, 65
 - CERN, 66
 - Tevatron, 67
- Parity violation in weak interactions, 187
- Particle interactions with matter, 74
 - Čerenkov radiation, 79
 - electromagnetic interactions, 77
 - hadronic e/h ratio, 81
 - hadronic interactions, 79
 - hadronic interaction length, 80
 - multiple scattering, 76
 - radiation length, 77
 - transition radiation, 79
- Particle travelling backwards in time, 151
- Particles, 5
- Physics at LEP
 - Beam energy measurement, 218
 - LEP1, 215
 - Z cross sections and forward–backward asymmetries, 219
 - Z line shape, 216
 - number of neutrinos, 217

LEP2, 221

- radiative return to the Z , 221
- W^+W^- production, 222
- luminosity measurement, 221
- outline, 215

Q

- Quadrupole magnet strength, 57
- Quantum chromodynamics (QCD)
 - $\alpha_s(Q^2)$
 - methods of determination, 268
 - coloured quark scattering
 - amplitude, 266
 - colour factors, 265
 - matrix elements, 264
 - Experimental tests of QCD gauge structure, 269
 - event shapes using thrust variable T , 270
 - introduction, 259
 - evidence for gluons, 260, 261
 - mathematical structure, 262–264
 - number of colours, 261
 - quark (parton) distribution functions (PDFs)
 - experimental determination, 271
 - scaling violations, 271
 - the gluon PDF, 271, 272
 - running coupling constants
 - $\alpha_{\text{QCD}}(Q^2)$, 268
 - $\alpha_s(Q^2)$, 268
 - QED example, 266, 267
- Quark model, 118
 - I -, U -, V -spins, 123
 - $J = 1$ vector mesons, 124
 - $I_3 = S = 0$ states: quark composition and decay modes, 125
 - octet–singlet mixing, 124
- antiquarks
 - charge conjugation, 119
- baryons, 118, 125
 - colour wavefunction, 125
 - proton spin–flavour wavefunction, 127
 - spin wavefunction, 126
 - SU(3) decuplet, 126
 - SU(3) octet, 126
 - SU(3)_{flavour} wavefunction, 126
- bottomonium
 - Υ discovery, 136
 - Υ resonances, 136
 - spectrum of states, 137
- charmonium
 - J/ψ J^{PC} , 133
 - J/ψ isospin, 132
 - J/ψ width, 133
 - $\psi(2S)$, $\psi(3S)$ states, 134
 - comparison with positronium, 134
 - decay systematics, 134

exotic hadrons, 137

- glueballs, 137
- tetraquarks, 137
- heavy quarks, 128
 - $Q\bar{Q}$ systems, 132
 - $b\bar{q}$ beauty mesons, 132
 - $c\bar{q}$ charmed mesons, 131
 - charm quark, 129
 - charmonium, 132
 - discovery of beauty, 130
 - discovery of charm, 129
- isospin, 119
 - $\Delta \rightarrow \pi$ -nucleon decays, 120
 - ud doublet, 119
 - baryonic (u, d) states with $I = \frac{3}{2}, \frac{1}{2}$, 120
 - fundamental assumptions, 119
 - pion $I = 1$ triplet, 120
- mesons, 118
- pseudoscalar mesons, 123
 - $I = S = 0$ states, 123
 - η , 123
 - η' , 123
 - π^0 , 123
- strange particles, 122
- SU(3)_{flavour}
 - fundamental representation, 122
 - SU(2) subgroups I -, U -, V -spins, 123
- Quark–parton model (QPM), 240
 - $\nu_e N$, $\bar{\nu}_e N$
 - $\sigma(\nu N)/E_\nu$ versus E_ν , 247
 - y distribution, quark spin- $\frac{1}{2}$, 247
 - nucleon antiquark/quark fraction, 247
 - nucleon quark/antiquark content, 246
 - $\nu_e N$, $\bar{\nu}_e N$ cross sections versus data, 246
 - $\nu_e N$, $\bar{\nu}_e N$ cross-section formulae, 246
 - y variable, 242
 - in terms of CMS scattering angle, 242
 - variable x
 - summary of interpretations, 241
 - definition of $q(x)$, $\bar{q}(x)$ distribution functions, 246
 - definition of the variable Q^2 , 241
 - definition of the variable x
 - in terms of energy transfer ν , 241
 - in terms of Lorentz invariants, 241
 - mass fraction m_q/M , 240
- hadron–hadron
 - Drell–Yan cross section, scaling variables, 273
 - Drell–Yan process, 273
 - formalism, 272
- key assumption, 240
- parton distribution functions (PDFs), 247
- validity in terms of Q^2 , 246

- R**
- Relativistic quantum field theory, 140
 - Relativistic quantum mechanics,
 - limitations of, 154
 - Relativistic quantum
 - mechanics/relativistic quantum field theory, discussion of, 148
 - RF acceleration
 - confined electromagnetic waves, 48
 - pill-box cavity, 49
 - Rotation in 3-dimensional space,
 - spin-matrix representation, 141
 - Rutherford scattering, 235
 - charge distribution
 - differential cross section, 237
 - form factor, 237, 238
 - Coulomb potential, 235
 - differential cross section, 237
 - matrix element, 236
 - matrix element for point-like particles, 236
 - scattering from nucleons, 239
- S**
- Scattering from a nucleon
 - kinematics, 239
 - elastic scattering condition, 239
 - SNO experiment
 - evidence for solar neutrino oscillations, 318
 - Solar Neutrino Oscillation (SNO)
 - experiment
 - Phase one, 317
 - Phase two, 318
 - Solar neutrinos, 315
 - Davis C_2Cl_4 experiment, 316
 - GALLEX experiment, 319
 - KamLAND $\bar{\nu}_e$ experiment
 - neutrinos from nuclear reactors, 322
 - survival probability oscillations, 323
 - minimum ($P_{\nu_e \rightarrow \nu_e}$) problem, 320
 - MSW matter effect, 320
 - MSW matter effect
 - E_{Res} resonance condition, 322
 - mathematical formalism, 320
 - mixing angle θ_m , 321
 - Standard Solar Model (SSM), 316
 - Spatial rotations, 23
 - conservation of angular momentum, 24
 - Euler angles, 25
 - rotation matrix, 25
 - Spin matrices, 2×2 complex matrix
 - vector space, 141
 - Spin- $\frac{1}{2}$ algebra, 115
 - commutation relations, 116
 - eigenvalues and eigenvectors, 116
 - three spin- $\frac{1}{2}$ states, 117
 - spin- $\frac{1}{2}$ states, 117
 - spin- $\frac{3}{2}$ states, 117
 - two spin- $\frac{1}{2}$ states, 116
 - Spin-matrix representation of spatial
 - vector, 142
 - Spinors, 142
 - dotted and undotted, 144
 - spatial rotation, 142
 - 4π -rotation experiment, 143
 - Weyl and Pauli, 144
 - Stochastic cooling, 66
 - Strange particles
 - discovery, 121
 - strangeness, 122
 - $SU(3)_{flavour}$, 122
 - Structure functions
 - Adler sum rule, 250
 - Gross-Llewellyn Smith sum rule, 250
 - Structure functions F_2 and F_3 for $\nu_e N$,
 - $\bar{\nu}_e N$, 248
 - Superconducting magnets
 - cable design, 60
 - energy stored, 61
 - LHC two-in-one design, 60
 - quench, 62
 - Rutherford cable, 60
 - Supersymmetry (SUSY), 368
 - particle spectrum, 369
 - sparticle pair production, 373
 - large-missing-energy signature, 374
 - sparticles
 - importance of the stop in searches, 378
 - main SM backgrounds, 375
 - QCD induced backgrounds, 376
 - Synchronicity condition, 49
 - Synchrotron radiation, 53
 - energy loss, 53
 - X-ray sources, 53
- T**
- Top-quark physics
 - $\sigma(t\bar{t})$ (combined CDF and D0), 231
 - b -tagging, 230
 - $t\bar{t}$ final-state topologies, 229
 - discovery, 229
 - top-quark mass, 231
 - Top-quark production
 - $t\bar{t}$ cross section vs \sqrt{s} , 367
 - Transitions and observables, 31
 - phase space, 32
- W**
- W and Z physics
 - W mass determination, 225
 - p_T^e distribution, 226
 - from $M_T(W \rightarrow \mu\nu\mu)$, 228
 - Jacobian peak, 226
 - W width determination, 229
 - $W \rightarrow \mu\nu\mu$ V-A test, 224
 - W^\pm charged-current weak quanta, 181
 - Weak interactions
 - CP violation in CKM matrix, 191
 - G_F and weak coupling g_w , 185
 - τ decay, 184
 - τ -decay matrix element, 185
 - CKM matrix CP violation phase, 192
 - CKM matrix parameterization, 191
 - CKM mixing matrix, 190
 - currents and fields, 187
 - current-current matrix element, 188
 - electromagnetic comparison, 188
 - Dirac bilinear covariants, 187
 - Fermi theory, 182
 - hadrons, 189
 - Cabibbo mixing (CKM matrix), 189
 - GIM mechanism, 189
 - heavy-quark decays, 192
 - K^0 decay problem, 189
 - leptons, 183
 - universality, 185
 - V-A interaction, 186
 - parity violation, 187
 - selection rules, 187
 - Weak neutral currents
 - discovery, 208
 - measurement of $\sin^2 \theta_W$, 208
 - Weinberg angle in electroweak
 - unification, 196
 - Weyl spinors
 - covariant eqns for $m \neq 0$, Dirac equation, 146
 - Lorentz-invariant Weyl equations, 146
 - metric tensor, 144
 - scalar products, 144
 - spin- $\frac{1}{2}$ massless particle, 146
 - transformation properties, 145
 - under space inversion, 145
- Z**
- Z^0 at LEP, 208
 - accelerator chain, 209
 - event classification, 210
 - OPAL detector, 209
 - physics analysis, 213
 - Z^0 neutral-current weak quantum, 181