Scottish Graduate Series

Einan Gardi
Nigel Glover
Aidan Robson *Editors*

LHC Phenomenology

Springer

LHC Phenomenology

Scottish Graduate Series

The Scottish Graduate Series is a long-standing series of graduate level texts proceeding from the Scottish Universities Summer Schools in Physics (SUSSP). SUSSP was established in 1960 to contribute to the dissemination of advanced knowledge in physics, and the formation of contacts among scientists from different countries through the setting up of a series of annual summer schools of the highest international standard. Each school is organized by its own committee which is responsible for inviting lecturers of international standing to contribute an in-depth lecture series on one aspect of the area being studied.

More information about this series at
http://www.springer.com/series/11662

Einan Gardi • Nigel Glover • Aidan Robson
Editors

LHC Phenomenology

Springer

Editors
Einan Gardi
Higgs Centre for Theoretical Physics
School of Physics and Astronomy
University of Edinburgh
Edinburgh, UK

Nigel Glover
Department of Physics
University of Durham
Science Laboratories
Durham, UK

Aidan Robson
School of Physics and Astronomy
University of Glasgow
Glasgow, UK

ISSN 2199-4617 ISSN 2155-2207 (electronic)
ISBN 978-3-319-05361-5 ISBN 978-3-319-05362-2 (eBook)
DOI 10.1007/978-3-319-05362-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014947539

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

SUSSP Proceedings

1	1960	Dispersion Relations
2	1961	Fluctuation, Relaxation and Resonance in Magnetic Systems
3	1962	Polarons and Excitons
4	1963	Strong Interactions and High Energy Physics
5	1964	Nuclear Structure and Electromagnetic Interactions
6	1965	Phonons in Perfect Lattices and in Lattices with Point Imperfections
7	1966	Particle Interactions at High Energy
8	1967	Methods in Solid State and Superfluid Theory
9	1968	Physics of Hot Plasmas
10	1969	Quantum Optics
11	1970	Hadronic Interactions of Electrons and Photons
12	1971	Atoms and Molecules in Astrophysics
13	1972	Electronic and Structural Properties of Amorphous Semiconductors
14	1973	Phenomenology of Particles at High Energy
15	1974	The Helium Liquids
16	1975	Non-linear Optics
17	1976	Fundamentals of Quark Models
18	1977	Nuclear Structure Physics
19	1978	Metal Non-metal Transitions in Disordered Solids
20	1979	Laser-Plasma Interactions 1
21	1980	Gauge Theories and Experiments at High Energy
22	1981	Magnetism in Solids
23	1982	Lasers: Physics, Systems and Techniques

SUSSP Proceedings (continued)

24	1982	Laser-Plasma Interactions 2
25	1983	Quantitative Electron Microscopy
26	1983	Statistical and Particle Physics
27	1984	Fundamental Forces
28	1985	Superstrings and Supergravity
29	1985	Laser-Plasma Interactions 3
30	1985	Synchrotron Radiation
31	1986	Localisation and Interaction
32	1987	Computational Physics
33	1987	Astrophysical and Laboratory Spectroscopy
34	1988	Optical Computing
35	1988	Laser-Plasma Interactions 4
36	1989	Physics of the Early Universe
37	1990	Pattern Recognition and Image Processing
38	1991	Physics of Nanostructures
39	1991	High Temperature Superconductivity
40	1992	Quantitative Microbeam Analysis
41	1992	Spatial Complexity in Optical Systems
42	1993	High Energy Phenomenology
43	1994	Determination of Geophysical Parameters from Space
44	1994	Simple Quantum Systems
45	1994	Laser-Plasma Interactions 5: Inertial Confinement Fusion
46	1995	General Relativity
47	1995	Laser Sources and Applications
48	1996	Generation and Application of High Microwaves Power
49	1997	Physical Processes in the Coastal Zone
50	1998	Semiconductor Quantum Optoelectronics
51	1998	Muon Science
52	1998	Advances in Lasers and Applications
53	1999	Soft and Fragile Matter
54	2000	The Restless Universe
55	2001	Heavy Flavour Physics
56	2002	Ultrafast Photonics
57	2003	Large Hadron Collider Phenomenology
58	2004	Hadron Physics
59	2004	Soft Condensed Matter Physics in Molecular and Cell Biology

SUSSP Proceedings (continued)

60	2005	Laser Plasma Interactions
61	2006	Neutrinos in Particle Physics, Astrophysics and Cosmology
62	2007	Extra-Solar Planets
63	2008	High Pressure Physics
64	2008	Advanced Techniques in Electron Spin Resonance
65	2009	LHC Physics
66	2010	Ultrafast Nonlinear Optics
67	2011	Quantum Information & Coherence
68	2011	Laser-Plasma Interactions and Applications
69	2012	LHC Phenomenology

The students of the school (alphabetical order)

Jose Didino GARCIA AGUILAR
Alexander BIEN
Armin BURGMEIER
Shreyashi CHAKDAR
Xiaoyong CHU
Daniel Souza COVACICH
Alina CZAJKA
Chiara DEBENEDETTI
Falko DULAT
Patrick ECKERT
Francisca GARAY
Johann Felix GRAF VON SODEN-FRAUNHOFEN
Tuomas HAPOLA
Maik HOESCHELE
Anna KOPP
Sabato LEO
Adrian LEWIS
Jeanette LORENZ
Clara Peset MARTIN
Hector Carranza MEJIA
Bernhard MISTLBERGER
Sebastian MORITZ
Huong NGUYEN
Stephen OGILVY
Tony POLL
Roxana RADU
Ipsita SAHA
Andrew ALTHEIMER
Marco-Andrea BUCHMANN
Adrian BUZATU
Nicola CHIAPOLINI
Pedro CIPRIANO
Robert CURRIE
Flavia DE ALMEIDA DIAS
Mirco DORIGO
Marc DUNSER
Laurent FORTHOMME
Rhorry GAULD
Nicolas GUTIERREZ
Tobias HARION
Zhen HU
Katharina KREPLIN
Christopher LESTER
Christian LINN
Haofei LUO
Marilyn MARX
Federico MELONI
Esben Tore MOLGAARD
Olaf NACKENHORST
Brendan O'BRIEN
Jacopo PAZZINI
Donnchadha QUILTY
Tom RAVENSCROFT
Paul SAIL

Ulises SALDANA-SALAZAR
Karoline Elfriede SELBACH
Suyog SHRESTHA
Ben SMART
Taylan TAKAN
Marzieh VAHABI
Chiu-Tien YU
Christoph ZIMMERMANN
Johannes SCHLENK
Varun SHARMA
Nikoloz SKHIRTLADZE
Jonathan STAHLMAN
Omran TRIFIS
Daniel Evangelho VIEIRA
Jia ZHOU
Max ZOLLER

Organising Committee

Professor C. Buttar	University of Glasgow	*Co-Director*
Professor F. Muheim	University of Edinburgh	*Co-Director*
Dr V. Martin	University of Edinburgh	*Secretary*
Dr S. Eisenhardt	University of Edinburgh	*Treasurer*
Prof. C.D. Froggatt	University of Glasgow	*Treasurer (deputising)*
Dr E. Gardi	University of Edinburgh	*Co-Editor*
Professor N. Glover	University of Durham	*Co-Editor*
Dr A. Robson	University of Glasgow	*Co-Editor*
Dr S. Scott	University of Glasgow	*Administration*

Preface

When we started preparing for the School in late 2010, the Large Hadron Collider (LHC) at CERN had been colliding beams for the first year with fast increasing luminosities. At that time we expected that, in 2011 and 2012, the LHC would provide large data samples such that the LHC experiments would take over the baton of particle physics at the frontier of energy and luminosity from the Tevatron and the B-factories. This made 2012 an opportune time to organise a Summer School with the topic of LHC results and phenomenology. However, the performance of the LHC exceeded all our expectations and, 1 month before the School was held in August 2012, the ATLAS and CMS experiments at CERN announced the discovery of a Higgs boson, making this result the hot topic of discussion at the School. This set the scene for a very successful 2 weeks. Furthermore, Peter Higgs himself came along to the School and gave a special lecture on how his ideas of electroweak symmetry breaking developed in the 1960s. He also highlighted the difference in the speed of communication now, with instant response available via email compared to taking several weeks via letter.

Following the pattern of many recent successful schools, we held the School in St. Andrews in August 2012, using the facilities of the Physics Department and accommodation at University Hall. This location is ideal for a School of this size (70 students and 10 lecturers and other staff) and character. The 70 participants came from 41 institutions, from Europe (Belgium, Denmark, Germany, Ireland, Italy, Norway, Poland, Spain, Switzerland and the UK), as well as Brazil, Chile, China, India, Mexico, Turkey and the USA.

The aim of the School was to equip young particle physicists with the basic tools to extract the maximum benefit from the various LHC experiments. This was achieved through a series of lectures providing an introduction to the theoretical and phenomenological framework of hadron collisions, and covering the recent results from the LHC. There were also lectures on the tools required by any particle physicist, theoretical or experimental, covering Monte Carlo models of interactions and statistical methods. The lectures were complemented by lively discussion classes covering the topics covered in the lectures and more widely.

With the European Particle Physics community preparing to review its strategy in Krakow in September 2012, the School held its own strategic review with a lively discussion amongst the lecturers and students on the future of particle physics in the Parliament Hall. The School hosted an outreach afternoon devoted to Schools that consisted of hands-on displays and experiments provided by the Edinburgh and Glasgow Particle Physics Experiment groups and a public lecture by Dr. Aidan Robson on the excitement of LHC physics. These proceedings provide a record of the lectures and will provide a valuable reference for those at the School and anyone wanting to develop a knowledge of the current status of particle physics.

Summer Schools are not just about science; they are about dialogue, discussion, meeting people from many backgrounds and forming lifelong friendships. The SUSSP has a tradition of hard work complemented by a lively social programme that brings together all the participants of the School. There were opportunities to sample Scottish culture, starting with a traditional ceilidh, through trips to local castles at Dunnottar and Glamis, a hill walk up Ben Vrackie, a memorable whisky tasting led by David Wishart, a very interesting walking tour of St. Andrews, a visit to Edinburgh with the opportunity to see the arts festival and the putting competition on the Himalayas. The School finished with a traditional Scottish banquet that included "haggies, neeps and tatties" accompanied by a piper. The after-dinner speech was given by Alan Walker (SUSSP Secretary/Treasurer), who presented a very humorous and entertaining account of the history and traditions of the SUSSP and there was a SUSSP first with participants performing a play, written by Jeff Richman, about the Higgs boson.

We would like to thank the lecturers for coming to St. Andrews and taking the time to assemble the lecturers and provide a written version for the proceedings. We also want to thanks very much all the students for coming from far and wide. We also thank our co-organisers Sean Benson, Stephan Eisenhardt, Colin Froggatt, Einan Gardi, Nigel Glover, Victoria Martin, Aidan Robson and Suzanne Scott from the universities of Edinburgh, Glasgow and the Institute for Particle Physics Phenomenology in Durham, for all their hard work. Everyone contributed to the lively and constructive atmosphere in the School, whether it was discussing the finer points of electroweak symmetry breaking, the future direction of particle physics or which is the best whisky.

We gratefully acknowledge the support of the UK Science and Technology Facilities Council (STFC), the European Science Foundation (ESF), the Scottish Universities Physics Alliance (SUPA), the Institute for Particle Physics Phenomenology in Durham, the Scottish Universities Summer Schools in Physics (particularly Alan Walker), the Institute of Physics, the Physics and Astronomy Departments of the Universities of Edinburgh and Glasgow without which the School would not have been possible. We are also grateful to University Hall and the Physics Department of St. Andrews for their generous hospitality.

Glasgow, UK — Prof Craig Buttar
Edinburgh, UK — Prof Franz Muheim
October 2013

Higgs in Hollywood

by

Jeffrey Richman

August 30, 2012

Cast of Characters: Bartender, Neutralino, Z boson, Higgs boson, Peter Higgs.

A Higgs boson, a Z, and a neutralino walk into a bar in Hollywood.

The bartender says, "What can I get you two?"

"You should have asked, 'What can I get you *three*?"' – said a tiny voice. It was the neutralino, a little bit of dark matter. The bartender looks all around the room. He sees nothing in the place where he thought he heard the voice coming from. Then he hears the little voice again.

"Hey, bartender! Over here! Give me a glass of your best whiskey. It's no fun being invisible when you have to account for a quarter of the energy density of the universe. I don't get *any respect!* In fact, I want to see some *real gratitude* for keeping those galactic rotation curves up to snuff! How would you feel if everyone ignored *you* for *all of recorded history!* Oh, I get *ill* every time I think about the photon – so much credit, and for what? The sun? Get serious!" He pauses, then slowly lifts his eyes upward. His arms outstretched, he says,"To be immortal and yet invisible: *why me?*"

"OK, OK," says the bartender. "I think I understand how it feels to be ignored by everyone." He pours a glass of whisky, puts it on the bar, and slides it expertly in the direction of the tiny voice. The whisky quickly drains out of the glass into thin air.

"What about you?" The bartender looks at the Z boson. "What will it be? You look like a vodka type to me."

"Oh . . . I'm feeling so FAT! Shouldn't have eaten that Goldstone boson. Couldn't resist breaking that symmetry. It was so . . . spontaneous. Before you could say 'broken generator' I had gained 90 GeV! I used to be massless, but I just can't get rid of this weight! Plus, my spin degrees of freedom will NEVER be the same. My longitudinal component is REALLY conspicuous. Is it showing? How embarrassing! Well, I am in Hollywood. Maybe I should get liposuction. Helicity zero! – that is so . . . rho meson! Oh . . . I could just decay right here on the spot! Goodbye cruel world!"

The bartender nods sympathetically and turns to the Higgs boson. "How about you? You look like a pretty massive particle yourself! Ha ha!"

The Higgs replies, "HELLOOOO! AS IF! I am, like, totally spinless at what you just said! All of my hidden valley-girl friends are going to say what a moron this bartender is."

"How DARE you call me that!" the bartender says. "I'm NOT a moron – I'm a moronino! Don't you KNOW THE DIFFERENCE! I bet you don't even know a Majorana spinor when you see one."

"WhatEVER", says the Higgs boson. But then she begins to cry.

"He's right – I am heavy. Sooo ... (she sighs) HEAVY. Right up there on the electroweak scale. Wow. Not only that, the vacuum just isn't what it used to be back in the early universe. The neighborhood has really gone downhill – I mean the Higgs potential has fallen to a new low. Really disgusting! Could someone PLEASE TURN DOWN THE VACUUM QUANTUM FLUCTUATIONS? This fine tuning is KILLING me. I can barely keep my mass under control. Could somebody come along and *stop* this madness? Pretty soon my GUTS are going to weigh as much as a gauge boson at the unification scale! Where are Weightwatchers when you need them?"

A distinguished looking gentleman enters the bar. The Higgs boson looks up.

"OMG, its Peter, my agent! He's soooo cute!"

She runs up to Peter and gives him a huge hug, which he is somewhat reluctant to accept. "Peter, come sit with your favorite client! You're looking good, Peter! How is life? Now that *my* talent has been discovered, I'll bet *you're* sitting pretty! Learning a bit of Swedish, ehhh? You are so lucky to be my agent. It's a win, win situation. You and me. Me and you."

She takes Peter's left hand, and holds it tightly in both of hers.

"Now Peter, I don't want to complain. The LHC job was good while it lasted, but, frankly, I need to move on. This is not all about your theories, as lovely as they are. Life is short. Very short. So do tell. Are there any exciting new jobs out there for me?"

"Being the most important manifestation of electroweak symmetry breaking isn't enough?" says Peter. "I thought that was a pretty good role for you."

"Peter ... Peter. I am not ungrateful for all that you have done. But, a particle like me, well, you know, ... needs an interacting job. And Hollywood should be just the place. I've heard there are interacting studios everywhere."

"Well, I'm not so sure they have any roles for you right now. The economy isn't so good. And I have to tell you. I have four other Higgs bosons now lined up as clients, and they have some absolutely *super* skills. One of them is CP odd! And I think that a couple of them can really charge up an audience."

The Higgs looks dejected. "But Peter, I know I can interact. True, its been a bit *weak* so far, but I've been thinking about increasing my coupling strength. I hear there are doctors here who can do anything."

The Z boson can't take it any more. "This is too much. Ms. Higgs boson here thinks she has all the talent. She thinks she can just change her coupling strength. Ridiculous! I am outta here."

The Z decays into two neutrinos, which immediately leave the bar without paying for the vodka. The neutralino drifts through the wall, out into the streets of Hollywood, and from there to the cosmos.

"Good-bye, Peter, says the Higgs. Here's to you. You predicted me! You said I would be great. You were *so* right!"

Suddenly, in a blinding flash of light, she decays into two photons.

The bartender says to himself, "At least she went out with a rare decay. She would have wanted it that way."

The bartender turns to Peter, who is the only one left in the bar. “I guess you’re the one who has to pay up. That’ll be 125 ± 1 dollars.”

Peter pulls out his wallet and pays.

“Thank you, Peter,” said the bartender, “you’re the best.”

THE END

Contents

Part III Tools

Contributors

Giacomo Cacciapaglia Université de Lyon, F-69622 Lyon, France
Université Lyon 1, Villeurbanne, France

CNRS/IN2P3, UMR5822, Institut de Physique Nucléaire de Lyon, Villeurbanne Cedex, France

Glen Cowan Physics Department, Royal Holloway, University of London, Egham, Surrey, UK

Tim Gershon Department of Physics, University of Warwick, Coventry, UK
European Organization for Nuclear Research (CERN), Geneva, Switzerland

Yuval Grossman Department of Physics, Newman Laboratory of Elementary Particle Physics, Cornell University, Ithaca, NY, USA

Sven Heinemeyer Instituto de Física de Cantabria (CSIC-UC), Santander, Spain

Marilyn Marx School of Physics and Astronomy, University of Manchester, Manchester, UK

Peter Mättig Bergische Universität Wuppertal, Wuppertal, Germany

William Murray RAL, Harwell Science and Innovation Campus, Harwell, UK

Jeffrey D. Richman Department of Physics, University of California, Santa Barbara, CA, USA

Michael H. Seymour School of Physics and Astronomy, University of Manchester, Manchester, UK

Part I
Theoretical Foundations

Higgs/Electroweak in the SM and the MSSM

Sven Heinemeyer

Abstract This lecture discusses the Higgs boson sectors of the SM and the MSSM, in particular in view of the recently discovered particle at $\sim$125.5 GeV. It also covers their connection to electroweak precision physics and the implications for the consistency tests of the respective models.

1 Introduction

A major goal of the particle physics program at the high energy frontier, currently being pursued at the CERN Large Hadron Collider (LHC), is to unravel the nature of electroweak symmetry breaking (EWSB). While the existence of the massive electroweak gauge bosons ($W^{\pm}, Z$), together with the successful description of their behaviour by non-abelian gauge theory, requires some form of EWSB to be present in nature, the underlying dynamics remained unknown for several decades. An appealing theoretical suggestion for such dynamics is the Higgs mechanism [1], which implies the existence of one or more Higgs bosons (depending on the specific model considered). Therefore, the search for a Higgs boson was considered a major cornerstone of the physics program at the LHC.

The spectacular discovery of a Higgs-like particle with a mass around $M_H \simeq 125.5\,\text{GeV}$, which has been announced by ATLAS [2] and CMS [3], marks a milestone in an effort that has been ongoing for almost half a century, and opens up a new era of particle physics. Both ATLAS and CMS reported a clear excess in the two photon channel, as well as in the $ZZ^{(*)}$ channel. The discovery is further corroborated, though not with high significance, by the $WW^{(*)}$ channel and by the final Tevatron results [4]. The combined sensitivity in each of the LHC experiments reaches more than 5σ.

Many theoretical models employing the Higgs mechanism in order to account for electroweak symmetry breaking have been studied in the literature, of which the most popular ones are the Standard Model (SM) [5] and the Minimal Supersymmetric Standard Model (MSSM) [6]. The newly discovered particle can be

S. Heinemeyer (✉)
Instituto de Física de Cantabria (CSIC-UC), Santander, Spain
e-mail: Sven.Heinemeyer@cern.ch

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_1

interpreted as the SM Higgs boson. The MSSM has a richer Higgs sector, containing two neutral $\mathscr{C}\mathscr{P}$-even, one neutral $\mathscr{C}\mathscr{P}$-odd, and two charged Higgs bosons. The newly discovered particle can also be interpreted as the light or the heavy $\mathscr{C}\mathscr{P}$-even state [7]. Among alternative theoretical models beyond the SM and the MSSM, the most prominent are the Two Higgs Doublet Model (THDM) [8], non-minimal supersymmetric extensions of the SM (e.g. extensions of the MSSM by an extra singlet superfield [9]), little Higgs models [10], and models with more than three spatial dimensions [11].

We will discuss the Higgs boson sector in the SM and the MSSM. This includes their agreement with the recently discovered particle around ∼125.5 GeV, their connection to electroweak precision physics, and the searches for the supersymmetric (SUSY) Higgs bosons at the LHC. While the LHC, after the discovery of a Higgs-like boson, will be able to measure some of its properties, a "cleaner" experimental environment, such as at the ILC, will be needed to measure all the Higgs boson characteristics [12–14].

2 The SM and the Higgs

2.1 Higgs: Why and How?

We start with looking at one of the most simple Lagrangians, the one of QED:

$$\mathscr{L}_{\rm QED} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \ . \tag{1}$$

Here D_μ denotes the covariant derivative

$$D_\mu = \partial_\mu + i\, e\, A_\mu \ . \tag{2}$$

ψ is the electron spinor, and A_μ is the photon vector field. The QED Lagrangian is invariant under the local $U(1)$ gauge symmetry,

$$\psi \to e^{-i\alpha(x)}\psi \ , \tag{3}$$

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha(x) \ . \tag{4}$$

Introduction of a mass term for the photon,

$$\mathscr{L}_{\text{photon mass}} = \tfrac{1}{2} m_A^2 A_\mu A^\mu \ , \tag{5}$$

is, however, not gauge-invariant. Applying Eq. (4) yields

$$\tfrac{1}{2} m_A^2 A_\mu A^\mu \to \tfrac{1}{2} m_A^2 \left[A_\mu A^\mu + \frac{2}{e} A^\mu \partial_\mu \alpha + \frac{1}{e^2} \partial_\mu \alpha\, \partial^\mu \alpha \right] \ . \tag{6}$$

A way out is the Higgs mechanism [1]. The simplest implementation uses one elementary complex scalar Higgs field Φ that has a vacuum expectation value (vev) v that is constant in space and time. The Lagrangian of the new Higgs field reads

$$\mathscr{L}_{\Phi} = \mathscr{L}_{\Phi,\text{kin}} + \mathscr{L}_{\Phi,\text{pot}} \tag{7}$$

with

$$\mathscr{L}_{\Phi,\text{kin}} = (D_\mu \Phi)^* \, (D^\mu \Phi) \, , \tag{8}$$

$$-\mathscr{L}_{\Phi,\text{pot}} = V(\Phi) = \mu^2 |\Phi|^2 + \lambda |\Phi|^4 \, . \tag{9}$$

Here λ has to be chosen positive to have a potential bounded from below. μ^2 can be either positive or negative, and we will see that $\mu^2 < 0$ yields the desired vev, as will be shown below. The complex scalar field Φ can be parametrized by two real scalar fields ϕ and η,

$$\Phi(x) = \frac{1}{\sqrt{2}} \phi(x) e^{i\eta(x)} \, , \tag{10}$$

yielding

$$V(\phi) = \frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \, . \tag{11}$$

Minimizing the potential one finds

$$\frac{\mathrm{d}V}{\mathrm{d}\phi}\Big|_{\phi=\phi_0} = \mu^2 \phi_0 + \lambda \phi_0^3 \stackrel{!}{=} 0 \, . \tag{12}$$

Only for $\mu^2 < 0$ does this yield the desired non-trivial solution

$$\phi_0 = \sqrt{\frac{-\mu^2}{\lambda}} \, (= \langle \phi \rangle =: v) \, . \tag{13}$$

The picture simplifies more by going to the "unitary gauge", $\alpha(x) = -\eta(x)/v$, which yields a real-valued Φ everywhere. The kinetic term now reads

$$(D_\mu \Phi)^* \, (D^\mu \Phi) \to \tfrac{1}{2}(\partial_\mu \phi)^2 + \tfrac{1}{2} e^2 q^2 \phi^2 A_\mu A^\mu \, , \tag{14}$$

where q is the charge of the Higgs field, which can now be expanded around its vev,

$$\phi(x) = v \; + \; H(x) \, . \tag{15}$$

The remaining degree of freedom, $H(x)$, is a real scalar boson, the Higgs boson. The Higgs boson mass and self-interactions are obtained by inserting Eq. (15) into the Lagrangian (neglecting a constant term),

$$-\mathscr{L}_{\text{Higgs}} = \tfrac{1}{2} m_H^2 H^2 + \frac{\kappa}{3!} H^3 + \frac{\xi}{4!} H^4 \,, \tag{16}$$

with

$$m_H^2 = 2\lambda v^2, \quad \kappa = 3\frac{m_H^2}{v}, \quad \xi = 3\frac{m_H^2}{v^2} \,. \tag{17}$$

Similarly, Eq. (15) can be inserted in Eq. (14), yielding (neglecting the kinetic term for ϕ),

$$\mathscr{L}_{\text{Higgs-photon}} = \tfrac{1}{2} m_A^2 A_\mu A^\mu + e^2 q^2 v H A_\mu A^\mu + \tfrac{1}{2} e^2 q^2 H^2 A_\mu A^\mu \tag{18}$$

where the second and third term describe the interaction between the photon and one or two Higgs bosons, respectively, and the first term is the photon mass,

$$m_A^2 = e^2 q^2 v^2 \,. \tag{19}$$

Another important feature can be observed: the coupling of the photon to the Higgs is proportional to its own mass squared.

Similarly, a gauge invariant Lagrangian can be defined to give mass to the chiral fermion $\psi = (\psi_L, \psi_R)^T$,

$$\mathscr{L}_{\text{fermion mass}} = y_\psi \psi_L^\dagger \, \Phi \, \psi_R + \text{c.c.} \,, \tag{20}$$

where y_ψ denotes the dimensionless Yukawa coupling. Inserting $\Phi(x) = (v + H(x))/\sqrt{2}$ one finds

$$\mathscr{L}_{\text{fermion mass}} = m_\psi \psi_L^\dagger \psi_R + \frac{m_\psi}{v} H \, \psi_L^\dagger \psi_R + \text{c.c.} \,, \tag{21}$$

with

$$m_\psi = y_\psi \frac{v}{\sqrt{2}} \,. \tag{22}$$

Again an important feature can be observed: by construction, the coupling of the fermion to the Higgs boson is proportional to its own mass m_ψ.

The "creation" of a mass term can be viewed from a different angle. The interaction of the gauge field or the fermion field with the scalar background field, i.e. the vev, shifts the masses of these fields from zero to non-zero values. This is shown graphically in Fig. 1 for the gauge boson (a) and the fermion (b) field.

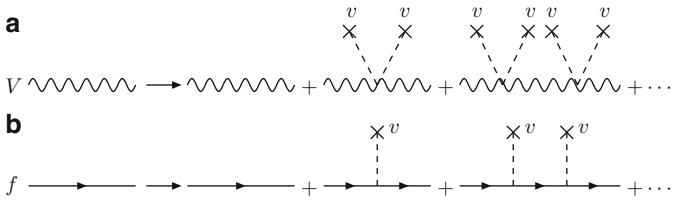

Fig. 1 Generation of a gauge boson mass (**a**) and a fermion mass (**b**) via the interaction with the vev of the Higgs field

The shift in the propagators reads (with p being the external momentum and $g = eq$ in Eq. (19)):

$$(a) \quad \frac{1}{p^2} \to \frac{1}{p^2} + \sum_{k=1}^{\infty} \frac{1}{p^2} \left[\left(\frac{gv}{2} \right) \frac{1}{p^2} \right]^k = \frac{1}{p^2 - m_V^2} \text{ with } m_V^2 = g^2 \frac{v^2}{4} , \quad (23)$$

$$(b) \quad \frac{1}{\not{p}} \to \frac{1}{\not{p}} + \sum_{k=1}^{\infty} \frac{1}{\not{p}} \left[\left(\frac{y_\psi v}{2} \right) \frac{1}{\not{p}} \right]^k = \frac{1}{\not{p} - m_\psi} \text{ with } m_\psi = y_\psi \frac{v}{\sqrt{2}} . \quad (24)$$

2.2 SM Higgs Theory

We now turn to the electroweak sector of the SM, which is described by the gauge symmetry $SU(2)_L \times U(1)_Y$. The bosonic part of the Lagrangian is given by

$$\mathscr{L}_{\text{bos}} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} + |D_\mu \Phi|^2 - V(\Phi), \quad (25)$$

$$V(\Phi) = \mu^2 |\Phi|^2 + \lambda |\Phi|^4 . \quad (26)$$

Φ is a complex scalar doublet with charges $(2, 1)$ under the SM gauge groups,

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} , \quad (27)$$

and the electric charge is given by $Q = T^3 + \frac{1}{2} Y$, where T^3 is the third component of the weak isospin. We furthermore have

$$D_\mu = \partial_\mu + i g \frac{\tau^a}{2} W_{\mu a} + i g' \frac{Y}{2} B_\mu , \quad (28)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu , \quad (29)$$

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g f^{abc} W_{\mu b} W_{\nu c} . \quad (30)$$

g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge couplings, respectively; τ^a are the Pauli matrices; and f^{abc} are the $SU(2)$ structure constants.

Choosing $\mu^2 < 0$, the minimum of the Higgs potential is found at

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v := \sqrt{\frac{-\mu^2}{\lambda}} \,. \tag{31}$$

$\Phi(x)$ can now be expressed through the vev, the Higgs boson and three Goldstone bosons $\phi_{1,2,3}$,

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\,\phi_2(x) \\ v + H(x) + i\,\phi_3(x) \end{pmatrix} \,. \tag{32}$$

Diagonalizing the mass matrices of the gauge bosons, one finds that the three massless Goldstone bosons are absorbed as longitudinal components of the three massive gauge bosons, $W_\mu^\pm, Z_\mu$, while the photon A_μ remains massless,

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right) \,, \tag{33}$$

$$Z_\mu = c_\mathrm{w} W_\mu^3 - s_\mathrm{w} B_\mu \,, \tag{34}$$

$$A_\mu = s_\mathrm{w} W_\mu^3 + c_\mathrm{w} B_\mu \,. \tag{35}$$

Here we have introduced the weak mixing angle $\theta_W = \arctan(g'/g)$, and $s_\mathrm{w} := \sin\theta_W$, $c_\mathrm{w} := \cos\theta_W$. The Higgs-gauge boson interaction Lagrangian reads,

$$\begin{aligned} \mathscr{L}_{\mathrm{Higgs-gauge}} = & \left[M_W^2 W_\mu^+ W^{-\mu} + \tfrac{1}{2} M_Z^2 Z_\mu Z^\mu \right] \left(1 + \frac{H}{v} \right)^2 \\ & - \tfrac{1}{2} M_H^2 H^2 - \frac{\kappa}{3!} H^3 - \frac{\xi}{4!} H^4 \,, \end{aligned} \tag{36}$$

with

$$M_W = \tfrac{1}{2} g v, \quad M_Z = \tfrac{1}{2} \sqrt{g^2 + g'^2}\, v, \tag{37}$$

$$(M_H^{\mathrm{SM}} :=)\; M_H = \sqrt{2\lambda}\, v, \quad \kappa = 3\frac{M_H^2}{v}, \quad \xi = 3\frac{M_H^2}{v^2} \,. \tag{38}$$

From the measurement of the gauge boson masses and couplings one finds $v \approx 246$ GeV. Furthermore the two massive gauge boson masses are related via

$$\frac{M_W}{M_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = c_\mathrm{w} \,. \tag{39}$$

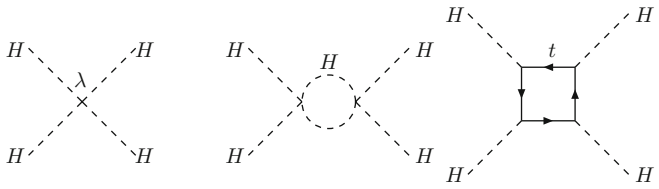

Fig. 2 Diagrams contributing to the evolution of the Higgs self-interaction λ at the tree level (*left*) and at the one-loop level (*middle* and *right*)

We now turn to the fermion masses, where we take the top- and bottom-quark masses as a representative example. The Higgs-fermion interaction Lagrangian reads

$$\mathscr{L}_{\text{Higgs-fermion}} = y_b Q_L^\dagger \, \Phi \, b_R \; + \; y_t Q_L^\dagger \, \Phi_c \, t_R + \; \text{h.c.} \tag{40}$$

$Q_L = (t_L, b_L)^T$ is the left-handed $SU(2)_L$ doublet. Going to the "unitary gauge" the Higgs field can be expressed as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} , \tag{41}$$

and it is obvious that this doublet can give masses only to the bottom(-type) fermion(s). A way out is the definition of

$$\Phi_c = i\sigma^2 \Phi^* \; = \; \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} , \tag{42}$$

which is employed to generate the top(-type) mass(es) in Eq. (40). Inserting Eqs. (41) and (42) into Eq. (40) yields

$$\mathscr{L}_{\text{Higgs-fermion}} = m_b \bar{b} b \left(1 + \frac{H}{v}\right) + m_t \bar{t} t \left(1 + \frac{H}{v}\right) \tag{43}$$

where we have used $\bar{\psi}\psi = \psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L$ and $m_b = y_b v/\sqrt{2}$, $m_t = y_t v/\sqrt{2}$.

The mass of the SM Higgs boson, M_H^{SM}, is in principle a free parameter in the model. However, it is possible to derive bounds on M_H^{SM} derived from theoretical considerations [15–17] and from experimental precision data. Here we review the first approach, while the latter one is followed in Sect. 2.5.

Evaluating loop diagrams as shown in the middle and right of Fig. 2 yields the renormalization group equation (RGE) for λ,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{3}{8\pi^2} \left[\lambda^2 + \lambda y_t^2 - y_t^4 + \frac{1}{16} \left(2g^4 + (g^2 + g'^2)^2\right) \right] , \tag{44}$$

with $t = \log(Q^2/v^2)$, where Q is the energy scale.

For large $M_H^2 \propto \lambda$, Eq. (44) reduces to

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{3}{8\pi^2}\lambda^2 \tag{45}$$

$$\Rightarrow \quad \lambda(Q^2) = \frac{\lambda(v^2)}{1 - \frac{3\lambda(v^2)}{8\pi^2}\log\left(\frac{Q^2}{v^2}\right)} \,. \tag{46}$$

For $\frac{3\lambda(v^2)}{8\pi^2}\log\left(\frac{Q^2}{v^2}\right) = 1$ one finds that λ diverges (it runs into the "Landau pole"). Requiring $\lambda(\Lambda) < \infty$ yields an upper bound on M_H^2 depending up to which scale Λ the Landau pole should be avoided,

$$\lambda(\Lambda) < \infty \;\Rightarrow\; M_H^2 \leq \frac{8\pi^2 v^2}{3\log\left(\frac{\Lambda^2}{v^2}\right)} \,. \tag{47}$$

For small $M_H^2 \propto \lambda$, on the other hand, Eq. (44) reduces to

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{3}{8\pi^2}\left[-y_t^4 + \frac{1}{16}\left(2g^4 + (g^2+g'^2)^2\right)\right] \tag{48}$$

$$\Rightarrow \quad \lambda(Q^2) = \lambda(v^2)\frac{3}{8\pi^2}\left[-y_t^4 + \frac{1}{16}\left(2g^4 + (g^2+g'^2)^2\right)\right]\log\left(\frac{Q^2}{v^2}\right) \,. \tag{49}$$

Demanding $V(v) < V(0)$, corresponding to $\lambda(\Lambda) > 0$, one finds a lower bound on M_H^2 depending on Λ,

$$\lambda(\Lambda) > 0 \;\Rightarrow\; M_H^2 \;>\; \frac{v^2}{4\pi^2}\left[-y_t^4 + \frac{1}{16}\left(2g^4 + (g^2+g'^2)^2\right)\right]\log\left(\frac{\Lambda^2}{v^2}\right) \,. \tag{50}$$

The combination of the upper bound in Eq. (47) and the lower bound in Eq. (50) on M_H is shown in Fig. 3. Requiring the validity of the SM up to the GUT scale yields a limit on the SM Higgs boson mass of 130 GeV $\lesssim M_H^{\mathrm{SM}} \lesssim$ 180 GeV.

2.3 *Predictions for a SM Higgs Boson at the LHC*

In order to efficiently search for the SM Higgs boson at the LHC precise predictions for the production cross sections and the decay branching ratios are necessary. To provide most up-to-date predictions in 2010 the "LHC Higgs Cross Section Working Group" [18] was founded. Two of the main results are shown in Fig. 4; see Refs. [19, 20] for an extensive list of references. The left plot shows the SM

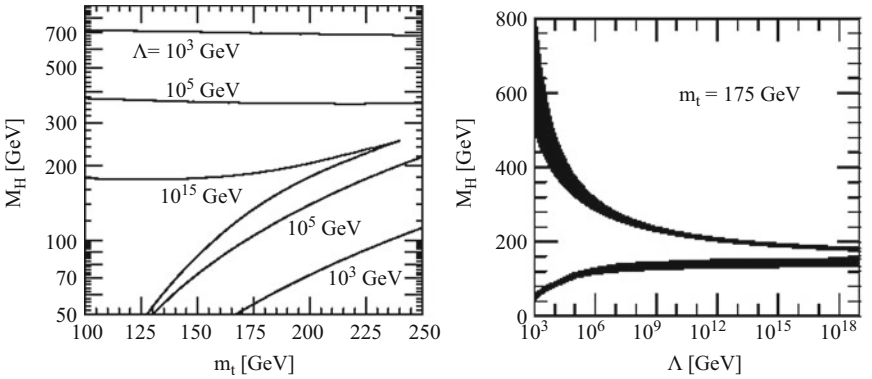

Fig. 3 Bounds on the mass of the Higgs boson in the SM. Λ denotes the energy scale up to which the model is valid [15–17]

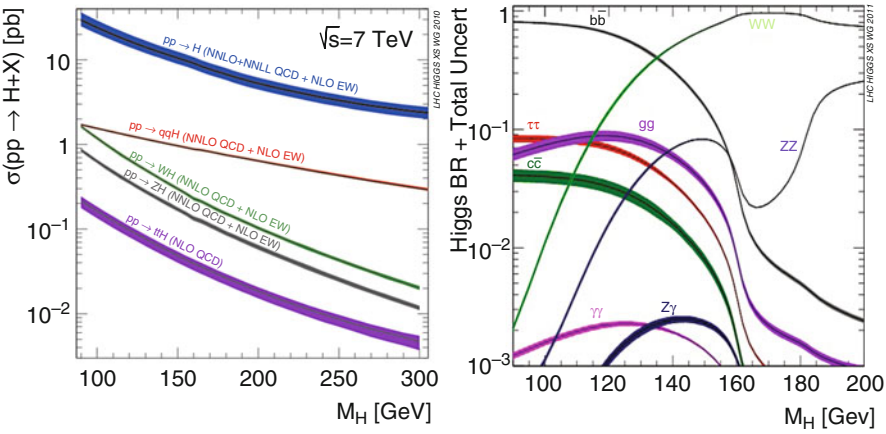

Fig. 4 Predictions for SM Higgs boson cross sections at the LHC with $\sqrt{s} = 7$ TeV (*left*) and the decay branching ratios (*right*) [19, 20]. The central lines show the predictions, while the colored bands indicate the theoretical uncertainty

theory predictions for the main production cross sections, where the colored bands indicate the theoretical uncertainties. (The same set of results is also available for $\sqrt{s} = 8$ TeV.) The right plot shows the branching ratios (BRs), again with the colored band indicating the theory uncertainty (see Ref. [21] for more details). Results of this type are constantly updated and refined by the Working Group.

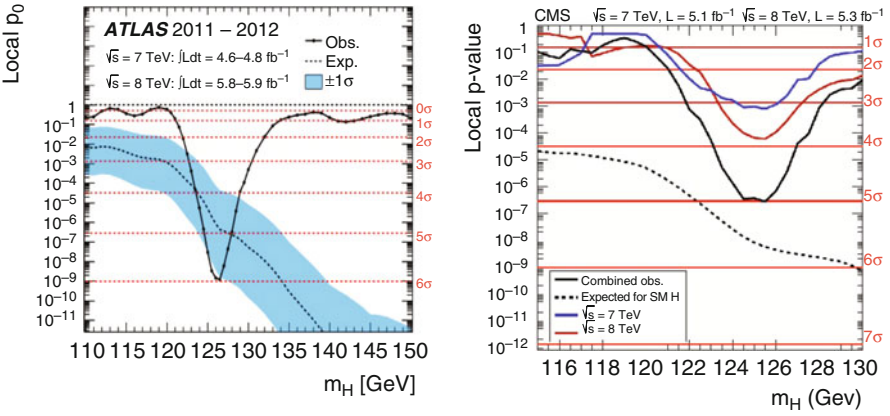

Fig. 5 p_0 values in the SM Higgs boson search (all channels combined) as presented by ATLAS (*left*) [2] and CMS (*right*) [3] on 4th of July 2012

2.4 *Discovery of an SM Higgs-Like Particle at the LHC*

On 4th of July 2012 both ATLAS [2] and CMS [3] announced the discovery of a new boson with a mass of ∼125.5 GeV. This discovery marks a milestone of an effort that has been ongoing for almost half a century and opens up a new era of particle physics. In Fig. 5 one can see the p_0 values of the search for the SM Higgs boson (with all search channels combined) as presented by ATLAS (left) and CMS (right) in July 2012. The p_0 value gives the probability that the experimental results observed can be caused by background only, i.e. in this case assuming the absence of a Higgs boson at each given mass. While the p_0 values are close to ∼0.5 for nearly all hypothetical Higgs boson masses (as would be expected for the absence of a Higgs boson), both experiments show a very low p_0 value of $p_0 \sim 10^{-6}$ around $M_H \sim 125.5$ GeV. This corresponds to the discovery of a new particle at the 5σ level by each experiment individually.

Another step in the analysis is a comparison of the measurement of production cross sections times branching ratios with the respective SM predictions, see Sect. 2.3. Two examples, using LHC data of about 5 fb^{-1} at 7 TeV and about 5 fb^{-1} at 8 TeV are shown in Fig. 6. Here ATLAS (left) [2] and CMS (right) [3] compare their experimental results with the SM prediction in various channels. It can be seen that all channels are, within the theoretical and experimental uncertainties, in agreement with the SM. However, it must be kept in mind that a measurement of the total width and thus of individual couplings is not possible at the LHC (see, e.g., Ref. [14] and references therein). Consequently, care must be taken in any coupling analysis. Recommendations of how these evaluations should be done using data from 2012 were given by the LHC Higgs Cross Section Working Group [22].

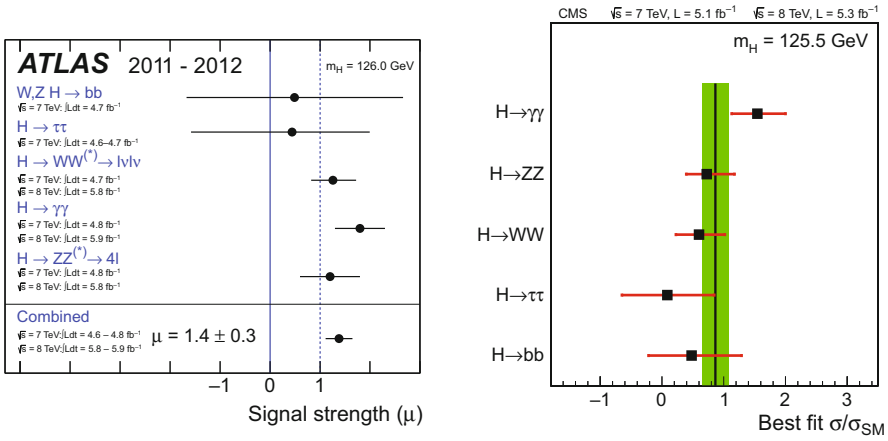

Fig. 6 Comparison of the measurement of production cross sections times branching ratios with the respective SM prediction from ATLAS [2] (*left*) and CMS [3] (*right*)

2.5 *Electroweak Precision Observables*

Within the SM the electroweak precision observables (EWPO) have been used in particular to constrain the SM Higgs-boson mass $M_H^{\rm SM}$, *before* the discovery of the new boson at ∼125.5 GeV. Originally the EWPO comprise over thousand measurements of "realistic observables" (with partially correlated uncertainties) such as cross sections, asymmetries, branching ratios etc. This huge set is reduced to 17 so-called "pseudo observables" by the LEP [23] and Tevatron [24] Electroweak working groups. The "pseudo observables" (again called EWPO in the following) comprise the W boson mass M_W, the width of the W boson, Γ_W, as well as various Z pole observables: the effective weak mixing angle, $\sin^2\theta_{\rm eff}$, Z decay widths to SM fermions, $\Gamma(Z \to f\bar{f})$, the invisible and total width, $\Gamma_{\rm inv}$ and Γ_Z, forward-backward and left-right asymmetries, $A_{\rm FB}^f$ and $A_{\rm LR}^f$, and the total hadronic cross section, $\sigma_{\rm had}^0$. The Z pole results including their combination are final [25]. Experimental progress in recent years has come from the Tevatron for M_W and m_t. (Also the error combination for M_W and Γ_W from the four LEP experiments has not yet been finalized due to not-yet-final analyses on the color-reconnection effects.)

The EWPO that give the strongest constraints on $M_H^{\rm SM}$ are M_W, $A_{\rm FB}^b$ and $A_{\rm LR}^e$. The value of $\sin^2\theta_{\rm eff}$ is extracted from a combination of various $A_{\rm FB}^f$ and $A_{\rm LR}^f$, where $A_{\rm FB}^b$ and $A_{\rm LR}^e$ give the dominant contribution.

The one-loop contributions to M_W can be decomposed as follows [26],

$$M_W^2\left(1-\frac{M_W^2}{M_Z^2}\right) = \frac{\pi\alpha}{\sqrt{2}G_F}(1+\Delta r)\,, \tag{51}$$

$$\Delta r_{1-\rm loop} = \Delta\alpha - \frac{c_{\rm w}^2}{s_{\rm w}^2}\Delta\rho + \Delta r_{\rm rem}(M_H^{\rm SM}). \tag{52}$$

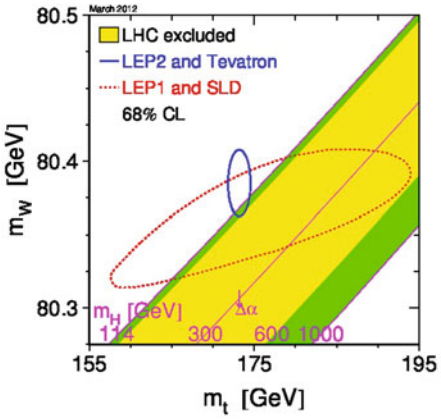

Fig. 7 Prediction for M_W in the SM as a function of m_t for the range $M_H^{\rm SM} = 114 \ldots 1{,}000$ GeV [23]. The yellow area for the range $M_H^{\rm SM} \sim 130 \ldots 600$ GeV is excluded by LHC searches for the SM Higgs boson. The prediction is compared with the present experimental results for M_W and m_t (at the 68 % CL) as well as with the indirect constraints obtained from EWPO

The first term, $\Delta\alpha$ contains large logarithmic contributions as $\log(M_Z/m_f)$ and amounts to ∼6 %. The second term contains the ρ parameter [27], with $\Delta\rho \sim m_t^2$. This term amounts to ∼3.3 %. The quantity $\Delta\rho$,

$$\Delta\rho = \frac{\Sigma^Z(0)}{M_Z^2} - \frac{\Sigma^W(0)}{M_W^2}, \tag{53}$$

parameterizes the leading universal corrections to the electroweak precision observables induced by the mass splitting between fields in an isospin doublet. $\Sigma^{Z,W}(0)$ denote the transverse parts of the unrenormalized Z and W boson self-energies at zero momentum transfer, respectively. The final term in Eq. (52) is $\Delta r_{\rm rem} \sim \log(M_H^{\rm SM}/M_W)$, and with a correction of size ∼1 % yields the constraints on $M_H^{\rm SM}$. The fact that the leading correction involving $M_H^{\rm SM}$ is logarithmic also applies to the other EWPO. Starting from two-loop order, terms $\sim (M_H^{\rm SM}/M_W)^2$ also appear. The SM prediction of M_W as a function of m_t for the range $M_H^{\rm SM} = 114 \ldots 1{,}000$ GeV is shown as the dark shaded (green) band in Fig. 7 [23], where an "intermediate region" of $M_H^{\rm SM} \sim 130 \ldots 600$ GeV as excluded by LHC SM Higgs searches is shown in yellow. The upper edge with $M_H^{\rm SM} = 114$ GeV corresponds to the (previous) lower limit on $M_H^{\rm SM}$ obtained at LEP [28]. The prediction is compared with the direct experimental result [23, 29],

$$M_W^{\rm exp} = 80.385 \pm 0.015 \text{ GeV}, \tag{54}$$

$$m_t^{\rm exp} = 173.2 \pm 0.9 \text{ GeV}, \tag{55}$$

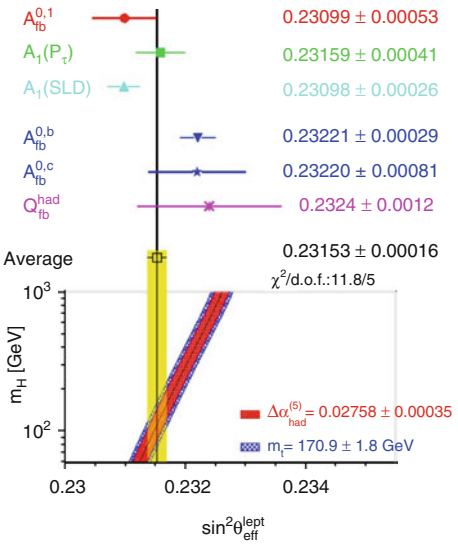

Fig. 8 Prediction for $\sin^2\theta_{\text{eff}}$ in the SM as a function of M_H^{SM} for $m_t = 170.9 \pm 1.8$ GeV and $\Delta\alpha_{\text{had}}^5 = 0.02758 \pm 0.00035$ [30]. The prediction is compared with the present experimental results for $\sin^2\theta_{\text{eff}}$ as averaged over several individual measurements

shown as the solid (blue) ellipse (at the 68 % CL) and with the indirect results for M_W and m_t as obtained from EWPO (dotted/red ellipse). The direct and indirect determination have significant overlap, representing a non-trivial success for the SM. Interpreting the newly discovered boson with a mass of ∼125.5 GeV as the SM Higgs boson, the plot shows agreement at the outer edge of the 68 % CL ellipse. However, it should be noted that the experimental value of M_W is somewhat higher than the region allowed by the LEP Higgs bounds: $M_H^{\text{SM}} \approx 60$ GeV is preferred as a central value by the measurement of M_W and m_t.

The effective weak mixing angle is evaluated from various asymmetries and other EWPO as shown in Fig. 8 [30] (no update taking into account more recent m_t measurements of this type of plot is available). The average determination yields $\sin^2\theta_{\text{eff}} = 0.23153 \pm 0.00016$ with a χ^2/d.o.f of 11.8/5, corresponding to a probability of 3.7 % [30]. The large χ^2 is driven by the two single most precise measurements, A_{LR}^e by SLD and A_{FB}^b by LEP, where the first (second) one prefers a value of $M_H^{\text{SM}} \sim 32(437)$ GeV (M. Grünewald, priv. communication). The two measurements differ by more than 3σ. The averaged value of $\sin^2\theta_{\text{eff}}$, as shown in Fig. 8, prefers $M_H^{\text{SM}} \sim 110$ GeV (M. Grünewald, priv. communication).

The indirect M_H^{SM} determination for several individual EWPO is given in Fig. 9. Shown in the left plot are the central values of M_H^{SM} and the one σ errors [23]. The dark shaded (green) vertical band indicates the combination of the various single measurements in the $1\,\sigma$ range. The vertical line shows the lower LEP bound for

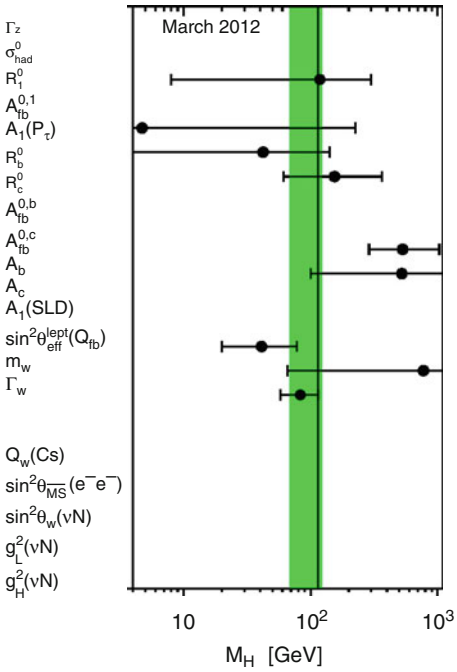

Fig. 9 Indirect constraints on $M_H^{\rm SM}$ from various EWPO. The central values for each observable are shown together with the one σ errors [23]. The *dark shaded (green)* vertical band indicates the combination of the various single measurements in the 1σ range. The *vertical line* shows the lower bound of $M_H^{\rm SM} \geq 114.4$ GeV obtained at LEP [28]

$M_H^{\rm SM}$ [28]. It can be seen that M_W, $A_{\rm LR}^e$ and $A_{\rm FB}^b$ give the most precise indirect $M_H^{\rm SM}$ determination, where only the latter one pulls the preferred $M_H^{\rm SM}$ value up, yielding a averaged value of [23]

$$M_H^{\rm SM} = 94^{+29}_{-24}\ {\rm GeV}\,, \tag{56}$$

which would be in agreement with the discovery of a new boson at $\sim$125.5 GeV. However, it is only the measurement of $A_{\rm FB}^b$ that yields the agreement of the SM with the new discovery.

In Fig. 10 [23] we show the result for the global fit to $M_H^{\rm SM}$ including all EWPO, but not including the direct search bounds from LEP, the Tevatron and the LHC. $\Delta\chi^2$ is shown as a function of $M_H^{\rm SM}$, yielding Eq. (56) as the best fit with an upper limit of 152 GeV at 95 % CL. The theory (intrinsic) uncertainty in the SM calculations (as evaluated with `TOPAZ0` [31] and `ZFITTER` [32]) are represented by the thickness of the blue band. The width of the parabola itself, on the other hand, is determined by the experimental precision of the measurements of the EWPO and the input parameters. Indicated as yellow areas are the $M_H^{\rm SM}$ values that are excluded by

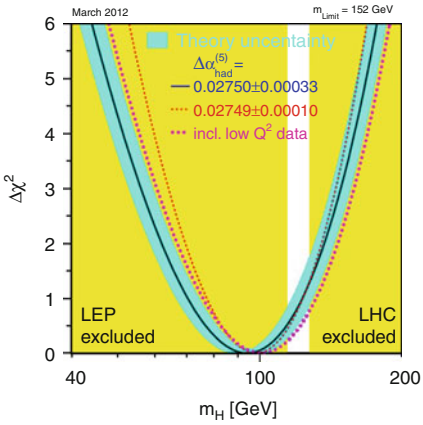

Fig. 10 $\Delta\chi^2$ curve derived from all EWPO measured at LEP, SLD, CDF and D0, as a function of $M_H^{\rm SM}$, assuming the SM to be the correct theory of nature, and not including the direct bounds on $M_H^{\rm SM}$ [23]

LEP and LHC searches, leaving only a small window of $M_H^{\rm SM} \sim 114 \ldots 130\,\text{GeV}$ open (reflecting that the plot was produced in March 2012). This window shrinks further taking into account the latest data from ATLAS [2] and CMS [3]. This plot demonstrates that a χ^2 penalty of $\sim$1 has to be paid to have $M_H^{\rm SM} \sim 125.5\,\text{GeV}$ with respect to the best fit value.

The current experimental uncertainties for the most relevant quantities, $\sin^2\theta_{\rm eff}$, M_W and m_t, can be substantially improved at the ILC and in particular with the GigaZ option [33–37]. It is expected that the leptonic weak effective mixing angle can be determined to 1.3×10^{-5}; for the W boson mass a precision of 7 MeV is expected. while for the top quark mass 0.1 GeV is anticipated from a precise determination of a well defined threshold mass. These improved accuracies will result in a substantially higher relative precision in the indirect determination of $M_H^{\rm SM}$, where with the GigaZ precision $\delta M_H^{\rm SM}/M_H^{\rm SM} \approx 16\,\%$ can be expected [30]. The comparison of the indirect $M_H^{\rm SM}$ determination with the direct measurement at the LHC [38, 39] and the ILC [40],

$$\delta M_H^{\rm SM,exp,LHC} \approx 200\,\text{MeV}, \tag{57}$$

$$\delta M_H^{\rm SM,exp,ILC} \approx 50\,\text{MeV}, \tag{58}$$

will constitute an important and profound consistency check of the model. This comparison will shed light on the basic theoretical components for generating the masses of the fundamental particles. On the other hand, an observed inconsistency would be a clear indication for the existence of a new physics scale.

3 The Higgs in Supersymmetry

3.1 Why SUSY?

Theories based on Supersymmetry (SUSY) [6] are widely considered as the theoretically most appealing extensions of the SM. They are consistent with the approximate unification of the gauge coupling constants at the GUT scale and provide a way to cancel the quadratic divergences in the Higgs sector hence stabilizing the huge hierarchy between the GUT and the Fermi scales. Furthermore, in SUSY theories the breaking of the electroweak symmetry is naturally induced at the Fermi scale, and the lightest supersymmetric particle can be neutral, weakly interacting and absolutely stable, providing therefore a natural solution for the dark matter problem.

The Minimal Supersymmetric Standard Model (MSSM) constitutes, hence its name, the minimal supersymmetric extension of the SM. The number of SUSY generators is $N = 1$, the smallest possible value. In order to keep anomaly cancellation, in contrast with to the SM a second Higgs doublet is needed [41]. All SM multiplets, including the two Higgs doublets, are extended to supersymmetric multiplets, resulting in scalar partners for quarks and leptons ("squarks" and "sleptons") and fermionic partners for the SM gauge boson and the Higgs bosons ("gauginos", "higgsinos" and "gluinos"). So far, the direct search for SUSY particles has not been successful. One can only set lower bounds of $\mathcal{O}(100\,\text{GeV})$ to $\mathcal{O}(1{,}000\,\text{GeV})$ on their masses [42].

3.2 The MSSM Higgs Sector

An excellent review on this subject is given in Ref. [43].

3.2.1 The Higgs Boson Sector at Tree-Level

In contrast with the Standard Model (SM), in the MSSM two Higgs doublets are required. The Higgs potential [44]

$$
\begin{aligned}
V = {} & m_1^2 |\mathscr{H}_1|^2 + m_2^2 |\mathscr{H}_2|^2 - m_{12}^2 (\epsilon_{ab} \mathscr{H}_1^a \mathscr{H}_2^b + \text{h.c.}) \\
& + \frac{1}{8}(g^2 + g'^2) \left[|\mathscr{H}_1|^2 - |\mathscr{H}_2|^2 \right]^2 + \frac{1}{2} g^2 |\mathscr{H}_1^\dagger \mathscr{H}_2|^2 \,, \qquad (59)
\end{aligned}
$$

contains m_1, m_2, m_{12} as soft SUSY breaking parameters; g, g' are the $SU(2)$ and $U(1)$ gauge couplings, and $\epsilon_{12} = -1$.

The doublet fields $\mathscr{H}_1$ and $\mathscr{H}_2$ are decomposed in the following way:

$$\mathscr{H}_1 = \begin{pmatrix} \mathscr{H}_1^0 \\ \mathscr{H}_1^- \end{pmatrix} = \begin{pmatrix} v_1 + \frac{1}{\sqrt{2}}(\phi_1^0 - i\chi_1^0) \\ -\phi_1^- \end{pmatrix},$$
$$\mathscr{H}_2 = \begin{pmatrix} \mathscr{H}_2^+ \\ \mathscr{H}_2^0 \end{pmatrix} = \begin{pmatrix} \phi_2^+ \\ v_2 + \frac{1}{\sqrt{2}}(\phi_2^0 + i\chi_2^0) \end{pmatrix}. \tag{60}$$

$\mathscr{H}_1$ gives mass to the down-type fermions, while $\mathscr{H}_2$ gives masses to the up-type fermions. The potential (59) can be described with the help of two independent parameters (besides g and g'): $\tan\beta = v_2/v_1$ and $M_A^2 = -m_{12}^2(\tan\beta + \cot\beta)$, where M_A is the mass of the $\mathscr{CP}$-odd Higgs boson A.

Which values can be expected for $\tan\beta$? One natural choice would be $\tan\beta \approx 1$, i.e. both vevs are about the same. On the other hand, one can argue that v_2 is responsible for the top quark mass, while v_1 gives rise to the bottom quark mass. Assuming that their mass differences comes largely from the vevs, while their Yukawa couplings could be about the same; the natural value for $\tan\beta$ would then be $\tan\beta \approx m_t/m_b$. Consequently, one can expect

$$1 \lesssim \tan\beta \lesssim 50\,. \tag{61}$$

The diagonalization of the bilinear part of the Higgs potential, i.e. of the Higgs mass matrices, is performed via the orthogonal transformations

$$\begin{pmatrix} H^0 \\ h^0 \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \phi_1^0 \\ \phi_2^0\,, \end{pmatrix} \tag{62}$$
$$\begin{pmatrix} G^0 \\ A^0 \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \chi_1^0 \\ \chi_2^0 \end{pmatrix}, \tag{63}$$
$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \phi_1^\pm \\ \phi_2^\pm \end{pmatrix}. \tag{64}$$

The mixing angle α is determined through

$$\alpha = \arctan\left[\frac{-(M_A^2 + M_Z^2)\sin\beta\cos\beta}{M_Z^2\cos^2\beta + M_A^2\sin^2\beta - m_{h,\text{tree}}^2}\right], \quad -\frac{\pi}{2} < \alpha < 0 \tag{65}$$

with $m_{h,\text{tree}}$ defined below in Eq. (69).

One gets the following Higgs spectrum:

$$\begin{aligned} \text{2 neutral bosons, } \mathscr{C}\mathscr{P} = +1 &: h, H \\ \text{1 neutral boson, } \mathscr{C}\mathscr{P} = -1 &: A \\ \text{2 charged bosons} &: H^+, H^- \\ \text{3 unphysical Goldstone bosons} &: G, G^+, G^-. \end{aligned} \tag{66}$$

At tree level the mass matrix of the neutral $\mathscr{C}\mathscr{P}$-even Higgs bosons is given in the ϕ_1-ϕ_2-basis in terms of M_Z, M_A, and $\tan\beta$ by

$$\begin{aligned} M_{\text{Higgs}}^{2,\text{tree}} &= \begin{pmatrix} m_{\phi_1}^2 & m_{\phi_1\phi_2}^2 \\ m_{\phi_1\phi_2}^2 & m_{\phi_2}^2 \end{pmatrix} \\ &= \begin{pmatrix} M_A^2 \sin^2\beta + M_Z^2 \cos^2\beta & -(M_A^2 + M_Z^2)\sin\beta\cos\beta \\ -(M_A^2 + M_Z^2)\sin\beta\cos\beta & M_A^2\cos^2\beta + M_Z^2\sin^2\beta \end{pmatrix}, \end{aligned} \tag{67}$$

which by diagonalization according to Eq. (62) yields the tree-level Higgs boson masses

$$M_{\text{Higgs}}^{2,\text{tree}} \xrightarrow{\alpha} \begin{pmatrix} m_{H,\text{tree}}^2 & 0 \\ 0 & m_{h,\text{tree}}^2 \end{pmatrix} \tag{68}$$

with

$$m_{H,h,\text{tree}}^2 = \tfrac{1}{2}\left[M_A^2 + M_Z^2 \pm \sqrt{(M_A^2 + M_Z^2)^2 - 4 M_Z^2 M_A^2 \cos^2 2\beta} \right]. \tag{69}$$

From this formula the famous tree-level bound

$$m_{h,\text{tree}} \leq \min\{M_A, M_Z\} \cdot |\cos 2\beta| \leq M_Z \tag{70}$$

can be obtained. The charged Higgs boson mass is given by

$$m_{H^\pm}^2 = M_A^2 + M_W^2 \,. \tag{71}$$

The masses of the gauge bosons are given in analogy to the SM:

$$M_W^2 = \frac{1}{2} g^2 (v_1^2 + v_2^2); \qquad M_Z^2 = \frac{1}{2}(g^2 + g'^2)(v_1^2 + v_2^2); \qquad M_\gamma = 0. \tag{72}$$

The couplings of the Higgs bosons are modified from the corresponding SM couplings already at tree-level. Some examples are

$$g_{hVV} = \sin(\beta-\alpha)\; g^{\mathrm{SM}}_{HVV}, \quad V = W^{\pm}, Z\,, \tag{73}$$

$$g_{HVV} = \cos(\beta-\alpha)\; g^{\mathrm{SM}}_{HVV}\,, \tag{74}$$

$$g_{hb\bar{b}}, g_{h\tau^+\tau^-} = -\frac{\sin\alpha}{\cos\beta}\; g^{\mathrm{SM}}_{Hb\bar{b},H\tau^+\tau^-}\,, \tag{75}$$

$$g_{ht\bar{t}} = \frac{\cos\alpha}{\sin\beta}\; g^{\mathrm{SM}}_{Ht\bar{t}}\,, \tag{76}$$

$$g_{Ab\bar{b}}, g_{A\tau^+\tau^-} = \gamma_5 \tan\beta\; g^{\mathrm{SM}}_{Hb\bar{b},H\tau^+\tau^-}\,. \tag{77}$$

The following can be observed: the couplings of the $\mathscr{CP}$-even Higgs boson to SM gauge bosons are always suppressed with respect to the SM coupling. However, if g^2_{hVV} is close to zero, g^2_{HVV} is close to $(g^{\mathrm{SM}}_{HVV})^2$ and vice versa, i.e. it is not possible to decouple both of them from the SM gauge bosons. The coupling of the h to down-type fermions can be suppressed *or enhanced* with respect to the SM value, depending on the size of $\sin\alpha/\cos\beta$. Especially for not too large values of M_A and large $\tan\beta$ one finds $|\sin\alpha/\cos\beta| \gg 1$, leading to a strong enhancement of this coupling. The same holds, in principle, for the coupling of the h to up-type fermions. However, for large parts of the MSSM parameter space the additional factor is found to be $|\cos\alpha/\sin\beta| < 1$. For the $\mathscr{CP}$-odd Higgs boson an additional factor $\tan\beta$ is found. According to Eq. (61) this can lead to a strongly enhanced coupling of the A boson to bottom quarks or τ leptons, resulting in new search strategies at the Tevatron and the LHC for the $\mathscr{CP}$-odd Higgs boson, see Sect. 3.3.

For $M_A \gtrsim 150\,\mathrm{GeV}$ the "decoupling limit" is reached. The couplings of the light Higgs boson become SM-like, i.e. the additional factors approach 1. The couplings of the heavy neutral Higgs bosons become similar, $g_{Axx} \approx g_{Hxx}$, and the masses of the heavy neutral and charged Higgs bosons fulfill $M_A \approx M_H \approx M_{H^\pm}$. As a consequence, search strategies for the A boson can also be applied to the H boson, and they are hard to disentangle at hadron colliders (see also Fig. 11 below).

3.2.2 The Scalar Quark Sector

Since the most relevant squarks for the MSSM Higgs boson sector are the $\tilde{t}$ and $\tilde{b}$ particles, here we explicitly list their mass matrices in the basis of the gauge eigenstates $\tilde{t}_L, \tilde{t}_R$ and $\tilde{b}_L, \tilde{b}_R$:

$$\mathscr{M}^2_{\tilde{t}} = \begin{pmatrix} M^2_{\tilde{t}_L} + m_t^2 + \cos 2\beta(\frac{1}{2} - \frac{2}{3}s^2_{\mathrm{w}})M^2_Z & m_t X_t \\ m_t X_t & M^2_{\tilde{t}_R} + m_t^2 + \frac{2}{3}\cos 2\beta s^2_{\mathrm{w}} M^2_Z \end{pmatrix}, \tag{78}$$

$$\mathscr{M}^2_{\tilde{b}} = \begin{pmatrix} M^2_{\tilde{b}_L} + m_b^2 + \cos 2\beta(-\frac{1}{2} + \frac{1}{3}s^2_{\mathrm{w}})M^2_Z & m_b X_b \\ m_b X_b & M^2_{\tilde{b}_R} + m_b^2 - \frac{1}{3}\cos 2\beta s^2_{\mathrm{w}} M^2_Z \end{pmatrix}. \tag{79}$$

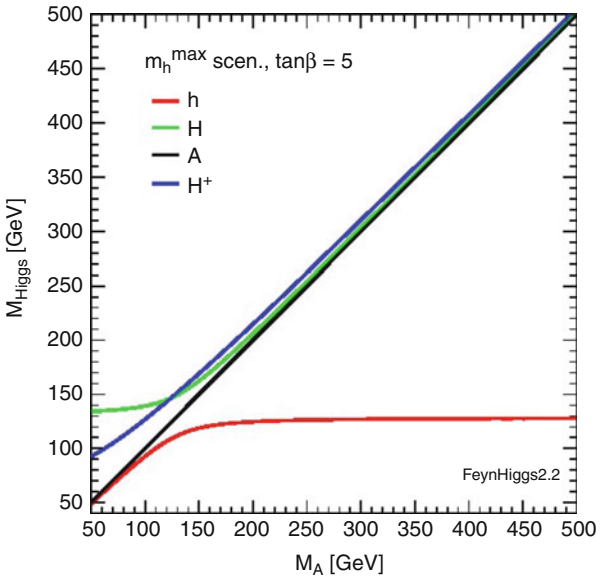

Fig. 11 The MSSM Higgs boson masses including higher-order corrections are shown as a function of M_A for $\tan\beta = 5$ in the $m_h^{\max}$ benchmark scenario [71] (Obtained with `FeynHiggs` [53, 65, 67, 68])

$M_{\tilde{t}_L}$, $M_{\tilde{t}_R}$, $M_{\tilde{b}_L}$ and $M_{\tilde{b}_R}$ are the (diagonal) soft SUSY-breaking parameters. We furthermore have

$$m_t X_t = m_t(A_t - \mu \cot\beta), \quad m_b X_b = m_b (A_b - \mu \tan\beta). \tag{80}$$

The soft SUSY-breaking parameters A_t and A_b denote the trilinear Higgs–stop and Higgs–sbottom coupling, and μ is the Higgs mixing parameter. $SU(2)$ gauge invariance requires the relation

$$M_{\tilde{t}_L} = M_{\tilde{b}_L}. \tag{81}$$

Diagonalizing $\mathscr{M}^2_{\tilde{t}}$ and $\mathscr{M}^2_{\tilde{b}}$ with the mixing angles $\theta_{\tilde{t}}$ and $\theta_{\tilde{b}}$, respectively, yields the physical $\tilde{t}$ and $\tilde{b}$ masses: $m_{\tilde{t}_1}$, $m_{\tilde{t}_2}$, $m_{\tilde{b}_1}$ and $m_{\tilde{b}_2}$.

3.2.3 Higher-Order Corrections to Higgs Boson Masses

A review about this subject can be found in Ref. [45]. In the Feynman diagrammatic (FD) approach the higher-order corrected $\mathscr{CP}$-even Higgs boson masses in the rMSSM are derived by finding the poles of the (h, H)-propagator matrix. The inverse of this matrix is given by

$$\left(\Delta_{\text{Higgs}}\right)^{-1} = -i \begin{pmatrix} p^2 - m^2_{H,\text{tree}} + \hat{\Sigma}_{HH}(p^2) & \hat{\Sigma}_{hH}(p^2) \\ \hat{\Sigma}_{hH}(p^2) & p^2 - m^2_{h,\text{tree}} + \hat{\Sigma}_{hh}(p^2) \end{pmatrix} . \quad (82)$$

Determining the poles of the matrix Δ_{Higgs} in Eq. (82) is equivalent to solving the equation

$$\left[p^2 - m^2_{h,\text{tree}} + \hat{\Sigma}_{hh}(p^2)\right]\left[p^2 - m^2_{H,\text{tree}} + \hat{\Sigma}_{HH}(p^2)\right] - \left[\hat{\Sigma}_{hH}(p^2)\right]^2 = 0 . \quad (83)$$

The very leading one-loop correction to M_h^2 is given by

$$\Delta M_h^2 = G_F m_t^4 \log\left(\frac{m_{\tilde{t}_1} m_{\tilde{t}_2}}{m_t^2}\right) , \quad (84)$$

where G_F denotes the Fermi constant. Equation (84) shows two important aspects: First, the leading loop corrections go with m_t^4, which is a "very large number". Consequently, the loop corrections can strongly affect M_h and push the mass beyond the reach of LEP [28,46] and into the mass regime of the newly discovered boson at $\sim$125.5 GeV. Second, the scalar fermion masses (in this case the scalar top masses) appear in the log entering the loop corrections (acting as a "cut-off" where the new physics enters). In this way the light Higgs boson mass depends on all other sectors via loop corrections. This dependence is particularly pronounced for the scalar top sector due to the large mass of the top quark.

The status of the available results for the self-energy contributions to Eq. (82) can be summarized as follows. For the one-loop part, the complete result is known within the MSSM [47–50]. The by far dominant one-loop contribution is the $\mathcal{O}(\alpha_t)$ term due to top and stop loops, see also Eq. (84), ($\alpha_t \equiv h_t^2/(4\pi)$, h_t being the superpotential top coupling). Computation of the two-loop effects is quite advanced and has now reached a stage such that all the contributions presumed to be dominant are known. They include the strong corrections, usually indicated as $\mathcal{O}(\alpha_t\alpha_s)$, and Yukawa corrections, $\mathcal{O}(\alpha_t^2)$, to the dominant one-loop $\mathcal{O}(\alpha_t)$ term, as well as the strong corrections to the bottom/sbottom one-loop $\mathcal{O}(\alpha_b)$ term ($\alpha_b \equiv h_b^2/(4\pi)$), i.e. the $\mathcal{O}(\alpha_b\alpha_s)$ contribution. The latter can be relevant for large values of $\tan\beta$. Currently, the $\mathcal{O}(\alpha_t\alpha_s)$ [51–55], $\mathcal{O}(\alpha_t^2)$ [51, 56, 57] and the $\mathcal{O}(\alpha_b\alpha_s)$ [58, 59] contributions to the self-energies are known for vanishing external momenta. In the (s)bottom corrections the all-order resummation of the $\tan\beta$-enhanced terms, $\mathcal{O}(\alpha_b(\alpha_s\tan\beta)^n)$ and $\mathcal{O}(\alpha_b(\alpha_t\tan\beta)^n)$, is also performed [60, 61]. The $\mathcal{O}(\alpha_t\alpha_b)$ and $\mathcal{O}(\alpha_b^2)$ corrections were presented in Ref. [62]. A "nearly full" two-loop effective potential calculation (including even the momentum dependence for the leading pieces and the leading three-loop corrections) has been published [63]. Most recently another leading three-loop calculation, valid for certain SUSY mass combinations, became available [64]. The remaining theoretical uncertainty on the lightest $\mathcal{CP}$-even Higgs boson mass has been estimated to be of $\sim$3 GeV [65, 66]. Taking the available loop corrections into account, the upper limit of M_h is shifted to [65],

$$M_h \leq 135\,\text{GeV} \quad (85)$$

(as obtained with the code FeynHiggs [53, 65, 67, 68]). This limit takes into account the experimental uncertainty for the top quark mass, see Eq. (55), as well as the intrinsic uncertainties from unknown higher-order corrections. Consequently, a Higgs boson with a mass of ∼125.5 GeV can naturally be explained by the MSSM. Either the light *or* the heavy $\mathscr{C}\mathscr{P}$-even Higgs boson can be interpreted as the newly discovered particle, which will be discussed in more detail in Sect. 3.4.

The charged Higgs boson mass is obtained by solving the equation

$$p^2 - m^2_{H^\pm} - \hat{\Sigma}_{H^-H^+}(p^2) = 0\,. \tag{86}$$

The charged Higgs boson self-energy is known at the one-loop level [69, 70].

3.3 MSSM Higgs Bosons at the LHC

The "decoupling limit" has been discussed for the tree-level couplings and masses of the MSSM Higgs bosons in Sect. 3.2.1. This limit also persists when radiative corrections are taken into account. The corresponding Higgs boson masses are shown in Fig. 11 for $\tan\beta = 5$ in the $m_h^{\max}$ benchmark scenario [71] obtained with FeynHiggs. For $M_A \gtrsim 180\,\text{GeV}$ the lightest Higgs boson mass approaches its upper limit (depending on the SUSY parameters), and the heavy Higgs boson masses are nearly degenerate. Furthermore, also the light Higgs boson couplings including loop corrections approach their SM-values. Consequently, for $M_A \gtrsim 180\,\text{GeV}$ an SM-like Higgs boson (below ∼135 GeV) can naturally be explained by the MSSM. On the other hand, deviations from a SM-like behavior can be described in the MSSM by moving from the full decoupling limit.

An example for the various production cross sections at the LHC is shown in Fig. 12 (for $\sqrt{s} = 14\,\text{TeV}$). For low masses the light Higgs cross sections are visible, and for $M_H \gtrsim 130\,\text{GeV}$ the heavy $\mathscr{C}\mathscr{P}$-even Higgs cross section is displayed, while the cross sections for the $\mathscr{C}\mathscr{P}$-odd A boson are given for the whole mass range. As discussed in Sect. 3.2.1 the g_{Abb} coupling is enhanced by $\tan\beta$ with respect to the corresponding SM value. Consequently, the $b\bar{b}A$ cross section is the largest or second largest cross section for all M_A, despite the relatively small value of $\tan\beta = 5$. For larger $\tan\beta$, see Eq. (61), this cross section can become even more dominant. Furthermore, the coupling of the heavy $\mathscr{C}\mathscr{P}$-even Higgs boson becomes very similar to the one of the A boson, and the two production cross sections, $b\bar{b}A$ and $b\bar{b}H$ are indistinguishable in the plot for $M_A > 200\,\text{GeV}$.

More precise results in the most important channels, $gg \to \phi$ and $b\bar{b} \to \phi$ ($\phi = h, H, A$) have been obtained by the LHC Higgs Cross Section Working Group [18], see also Refs. [19,20] and references therein. Most recently a new code, SusHi [73] for the $gg \to \phi$ production mode including the full MSSM one-loop contributions as well as higher-order SM and MSSM corrections has been presented, see Ref. [74] for more details.

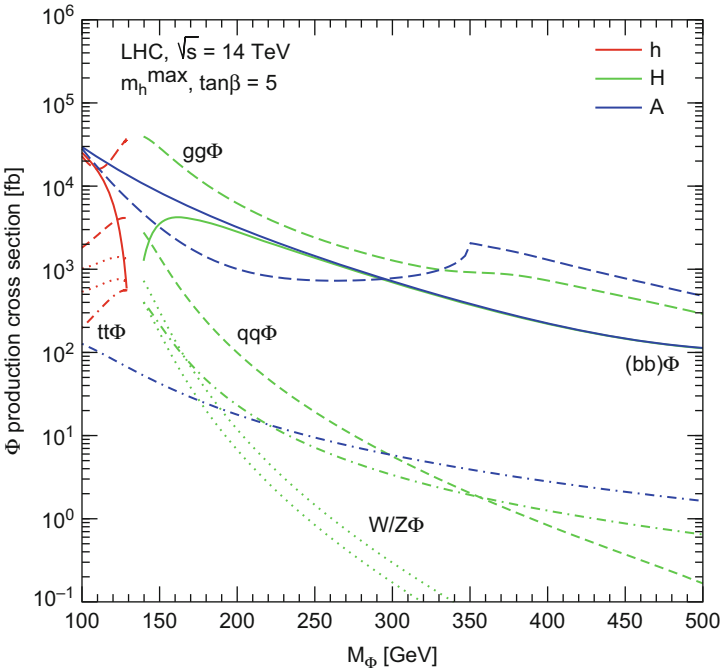

Fig. 12 Overview of the various neutral Higgs boson production cross sections at the LHC shown as a function of M_A for $\tan\beta = 5$ in the $m_h^{\max}$ scenario (Taken from Ref. [72], where the original references can be found)

Following the above discussion, the main search channel for heavy Higgs bosons at the LHC for $M_A \gtrsim 200\,\text{GeV}$ is the production in association with bottom quarks and the subsequent decay to tau leptons, $b\bar{b} \to b\bar{b}\ H/A \to b\bar{b}\ \tau^+\tau^-$. For heavy supersymmetric particles, with masses far above the Higgs boson mass scale, one has for the production and decay of the A boson [75]

$$\sigma(b\bar{b}A) \times \text{BR}(A \to b\bar{b}) \simeq \sigma(b\bar{b}H)_{\text{SM}} \frac{\tan^2\beta}{(1+\Delta_b)^2} \times \frac{9}{(1+\Delta_b)^2+9}\,, \tag{87}$$

$$\sigma(gg, b\bar{b} \to A) \times \text{BR}(A \to \tau^+\tau^-) \simeq \sigma(gg, b\bar{b} \to H)_{\text{SM}} \frac{\tan^2\beta}{(1+\Delta_b)^2+9}\,, \tag{88}$$

where $\sigma(b\bar{b}H)_{\text{SM}}$ and $\sigma(gg, b\bar{b} \to H)_{\text{SM}}$ denote the values of the corresponding SM Higgs boson production cross sections for $M_H^{\text{SM}} = M_A$. The leading contributions to Δ_b are given by [60]

$$\Delta_b \approx \frac{2\alpha_s}{3\pi} m_{\tilde{g}}\, \mu\, \tan\beta \times I(m_{\tilde{b}_1}, m_{\tilde{b}_2}, m_{\tilde{g}}) + \frac{\alpha_t}{4\pi} A_t\, \mu\, \tan\beta \times I(m_{\tilde{t}_1}, m_{\tilde{t}_2}, |\mu|)\,, \tag{89}$$

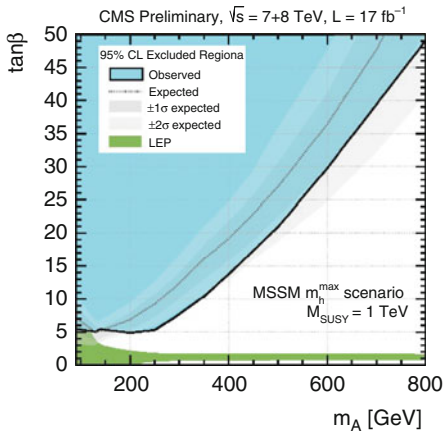

Fig. 13 The 95 % CL exclusion regions (i.e. the upper bound of the "LHC wedge" region) for the heavy neutral Higgs bosons in the channel $pp \to H/A \to \tau^+\tau^-(+X)$, obtained by CMS including $\sqrt{s} = 7, 8\,\text{TeV}$ data [76]

where the function I arises from the one-loop vertex diagrams and scales as $I(a, b, c) \sim 1/\max(a^2, b^2, c^2)$. Here $m_{\tilde{g}}$ is the gluino mass, and μ is the Higgs mixing parameter. As a consequence, the $b\bar{b}$ production rate depends sensitively on $\Delta_b \propto \mu \tan\beta$ because of the factor $1/(1 + \Delta_b)^2$, while this leading dependence on Δ_b cancels out in the $\tau^+\tau^-$ production rate. The formulas above apply, within a good approximation, also to the heavy $\mathscr{CP}$-even Higgs boson in the large $\tan\beta$ regime. Therefore, the production and decay rates of H are governed by similar formulas as the ones given above, leading to an approximate enhancement by a factor 2 of the production rates with respect to the ones that would be obtained in the case of the single production of the $\mathscr{CP}$-odd Higgs boson as given in Eqs. (87) and (88).

Of particular interest is the "LHC wedge" region, i.e. the region in which only the light $\mathscr{CP}$-even MSSM Higgs boson, but none of the heavy MSSM Higgs bosons can be detected at the LHC. It appears for $M_A \gtrsim 200\,\text{GeV}$ at intermediate $\tan\beta$ and widens to larger $\tan\beta$ values for larger M_A. Consequently, in the "LHC wedge" only a SM-like light Higgs boson can be discovered at the LHC, and part of the LHC wedge (depending on the explicit choice of SUSY parameters) can be in agreement with $M_h \sim 125.5\,\text{GeV}$. This region, bounded from above by the 95 % CL exclusion contours for the heavy neutral MSSM Higgs bosons can be seen in Fig. 13 [76]. Here it should be kept in mind that the actual position of the exclusion contour depends on Δ_b and thus on the sign and the size of μ as discussed above.

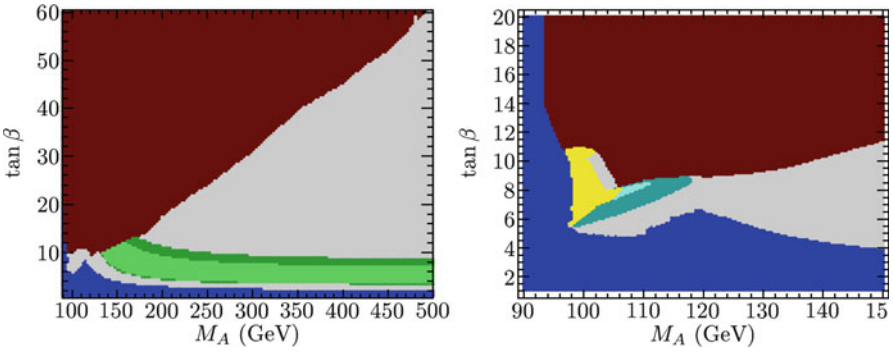

Fig. 14 *Left*: M_A-$\tan\beta$ plane in the $m_h^{\max}$ scenario; the *green shaded area* yields $M_h \sim 125.5$ GeV, the *brown area* is excluded by LHC heavy MSSM Higgs boson searches, the *blue area* is excluded by LEP Higgs searches. *Right*: M_A-$\tan\beta$ plane with $M_{\rm SUSY} = \mu = 1$ TeV, $X_t = 2.3$ TeV; the *yellow area* yields $M_H \sim 125.5$ GeV with an SM-like heavy $\mathcal{CP}$-even Higgs boson, *brown and blue areas* are excluded by LHC and LEP Higgs searches, respectively [7]

3.4 Agreement of the MSSM Higgs Sector with a Higgs at ∼125.5 GeV

Many investigations have been performed analyzing the agreement of the MSSM with a Higgs boson at ∼125.5 GeV. In a first step only the mass information can be used to test the model, while in a second step also the rate information of the various Higgs search channels can be taken into account. Here we briefly review the first MSSM results [7] that were published after the first ATLAS/CMS announcement in December 2012 [77] (see Refs. [78, 79] for updates of these results, including rate analyses, and for an extensive list of references).

In the left plot of Fig. 14 [7] the M_A-$\tan\beta$ plane in the $m_h^{\max}$ benchmark scenario [71] is shown, where the areas in light and dark green yield a mass for the light $\mathcal{CP}$-even Higgs around ∼125.5 GeV. The brown area is excluded by LHC heavy MSSM Higgs boson searches in the $H/A \to \tau\tau$ channel (although not by the latest results as presented in Ref. [76]): the blue area is excluded by LEP Higgs searches [28,46]. (The limits have been obtained with `HiggsBounds` [80] version 3.5.0-beta.) Since the $m_h^{\max}$ scenario maximizes the light $\mathcal{CP}$-even Higgs boson mass it is possible to extract lower (one parameter) limits on M_A and $\tan\beta$ from the edges of the green band. By choosing the parameters entering via radiative corrections such that those corrections yield a maximum upward shift to M_h, the lower bounds on M_A and $\tan\beta$ that can be obtained are general in the sense that they (approximately) hold for *any* values of the other parameters. To address the (small) residual $M_{\rm SUSY}(:= M_{\tilde{t}_L} = M_{\tilde{t}_R} = M_{\tilde{b}_R})$ dependence of the lower bounds on M_A and $\tan\beta$, limits have been extracted for the three different values $M_{\rm SUSY} = \{0.5, 1, 2\}$ TeV, see Table 1. For comparison, the previous limits derived from the LEP Higgs searches [46] are also shown, i.e. before the incorporation of

Table 1 Lower limits on the MSSM Higgs sector tree-level parameters M_A ($M_{H^\pm}$) and $\tan\beta$ obtained with and without the assumed Higgs signal of $M_h \sim 125.5$ GeV. The mass limits have been rounded to 1 GeV [7]

	Limits without $M_h \sim 125$ GeV			Limits with $M_h \sim 125$ GeV		
$M_{\rm SUSY}$ (GeV)	$\tan\beta$	M_A (GeV)	$M_{H^\pm}$ (GeV)	$\tan\beta$	M_A (GeV)	$M_{H^\pm}$ (GeV)
500	2.7	95	123	4.5	140	161
1,000	2.2	95	123	3.2	133	155
2,000	2.0	95	123	2.9	130	152

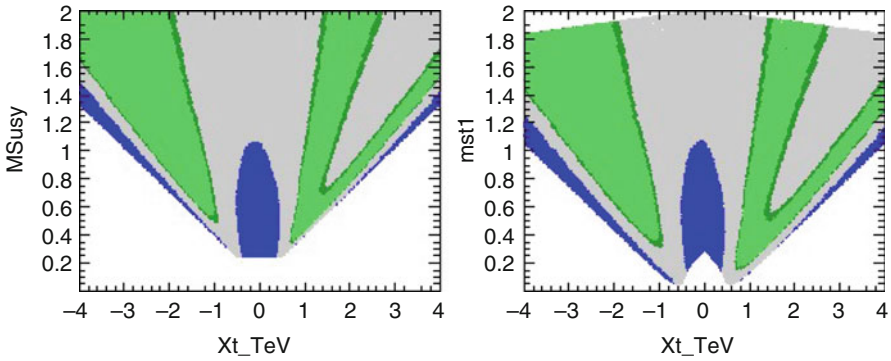

Fig. 15 Scalar top masses in the $m_h^{\rm max}$ scenario (with $M_{\rm SUSY}$ and X_t free) that yield $M_h \sim 125.5$ GeV (*green area*), LEP excluded regions are shown in *blue*. *Left*: X_t-$M_{\rm SUSY}$ plane, *right*: X_t-$m_{\tilde{t}_1}$ plane [7]

the new LHC results reported in Ref. [77]. The bounds on M_A translate directly into lower limits on $M_{H^\pm}$, which are also given in the table. A phenomenological consequence of the bound $M_{H^\pm} \gtrsim 155$ GeV (for $M_{\rm SUSY} = 1$ TeV) is that it would leave only a very small kinematic window open for the possibility that MSSM charged Higgs bosons are produced in the decay of top quarks.

It is also possible to investigate what can be inferred from the assumed Higgs signal about higher-order corrections in the Higgs sector. Similarly to the previous case, one can obtain an absolute lower limit on the stop mass scale $M_{\rm SUSY}$ by considering the maximal tree-level contribution to M_h. The resulting constraints for $M_{\rm SUSY}$ and X_t, obtained in the decoupling limit for $M_A = 1$ TeV and $\tan\beta = 20$, are shown in the left plot of Fig. 15 [7] with the same colour coding as before. Several favoured branches develop in this plane, centred around $X_t \sim -1.5 M_{\rm SUSY}$, $X_t \sim 1.2 M_{\rm SUSY}$, and $X_t \sim 2.5 M_{\rm SUSY}$. The minimal allowed stop mass scale is $M_{\rm SUSY} \sim 300$ GeV with positive X_t and $M_{\rm SUSY} \sim 500$ GeV for negative X_t. The results on the stop sector can also be interpreted as a lower limit on the mass $m_{\tilde{t}_1}$ of the lightest stop squark. This is shown in the right plot of Fig. 15. Interpreting the newly observed particle as the light $\mathscr{CP}$-even Higgs one obtains the lower bounds $m_{\tilde{t}_1} > 100$ GeV ($X_t > 0$) and $m_{\tilde{t}_1} > 250$ GeV ($X_t < 0$).

Finally, in the right plot of Fig. 14 [7] it is demonstrated that also the heavy $\mathscr{CP}$-even Higgs can be interpreted as the newly discovered particle at ∼125.5 GeV. The M_A-tan β plane is shown for $M_{\rm SUSY} = \mu = 1\,\text{TeV}$ and $X_t = 2.3\,\text{TeV}$. As before the blue region is LEP excluded, and the brown area indicates the bounds from $H/A \to \tau\tau$ searches. This area grows substantially when the latest results from Ref. [76] are taken into account. However, the scenario cannot be excluded, since no dedicated study for this part of the MSSM parameter space exists, and the limits from the $m_h^{\rm max}$ scenario cannot be carried over in a naive way. Requiring in addition that the production and decay rates into $\gamma\gamma$ and vector bosons are at least 90 % of the corresponding SM rates, a small allowed region is found (yellow). In this region, enhancements of the rate of up to a factor of three as compared to the SM rate are possible. In this kind of scenario M_h is found *below* the SM LEP limit of 114.4 GeV [28] exhibiting reduced couplings to gauge bosons so that the limits from the LEP searches for non-SM like Higgs bosons are respected [46].

3.5 Electroweak Precision Observables

Within the MSSM one can attempt to fit the unknown parameters to the existing experimental data, in a similar fashion as discussed in Sect. 2.5. However, fits within the MSSM differ from the SM fit in various ways. First, the number of free parameters is substantially larger in the MSSM, even restricting to GUT based models as discussed below. On the other hand, more observables can be taken into account, providing extra constraints on the fit. Within the MSSM the additional observables included are the anomalous magnetic moment of the muon $(g-2)_\mu$, B-physics observables such as BR$(b \to s\gamma)$, BR$(B_s \to \mu\mu)$, or BR$(B_u \to \tau\nu_\tau)$, and the relic density of cold dark matter (CDM), which can be provided by the lightest SUSY particle, the neutralino. These additional constraints would either have a minor impact on the best-fit regions or cannot be accommodated in the SM. Finally, as discussed in the previous subsections, whereas the light Higgs boson mass is a free parameter in the SM, it is a function of the other parameters in the MSSM. In this way, for example, the masses of the scalar tops and bottoms enter not only directly into the prediction of the various observables, but also indirectly via their impact on M_h.

Within the MSSM the dominant SUSY correction to electroweak precision observables arises from the scalar top and bottom contribution to the ρ parameter, see Eq. (53). Generically one finds $\Delta\rho^{\rm SUSY} > 0$, leading, for instance, to an upward shift in the prediction of M_W with respect to the SM prediction. The experimental result and the theory prediction of the SM and the MSSM for M_W are compared in Fig. 16 (updated from Ref. [81], see also Ref. [82]). The predictions within the two models give rise to two bands in the m_t–M_W plane, one for the SM and one for the MSSM prediction, where in each band either the SM Higgs boson or the light $\mathscr{CP}$-even MSSM Higgs boson is interpreted as the newly discovered particle at ∼125.5 GeV. Consequently, the respective Higgs boson masses are restricted to

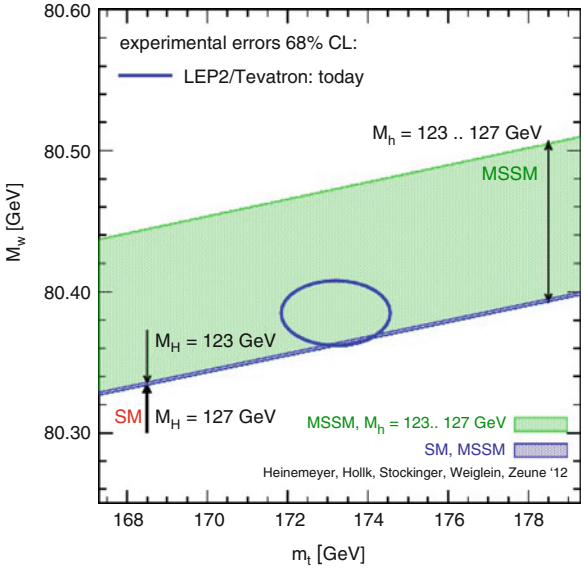

Fig. 16 Prediction for M_W in the MSSM and the SM (see text) as a function of m_t in comparison with the present experimental results for M_W and m_t (Updated from Ref. [81], see Refs. [66, 82] for more details)

be in the interval 123 ... 127 GeV. The SM region, shown as dark-shaded (blue) completely overlaps with the lower M_W region of the MSSM band, shown as light shaded (green). The full MSSM region, i.e. the light shaded (green) and the dark-shaded (blue) areas are obtained from scattering the relevant parameters independently [81, 82]. The decoupling limit with SUSY masses of $\mathcal{O}(2\,\text{TeV})$ yields the lower edge of the dark-shaded (blue) area. The current 68 and 95 % CL experimental results for m_t, Eq. (55), and M_W, Eq. (54), are also indicated in the plot. As can be seen from Fig. 16, the current experimental 68 % CL region for m_t and M_W exhibits a slight preference of the MSSM over the SM. This example indicates that the experimental measurement of M_W in combination with m_t prefers, within the MSSM, not too heavy SUSY mass scales.

As mentioned above, in order to restrict the number of free parameters in the MSSM one can resort to GUT based models. Most fits have been performed in the Constrained MSSM (CMSSM), in which the input scalar masses m_0, gaugino masses $m_{1/2}$ and soft trilinear parameters A_0 are each universal at the GUT scale, $M_{\text{GUT}} \approx 2 \times 10^{16}$ GeV, and in the Non-universal Higgs mass model (NUHM1), in which a common SUSY-breaking contribution to the Higgs masses is allowed to be non-universal (see Ref. [83] for detailed definitions). The results for the fits of M_h in the CMSSM and the NUHM1 are shown in Fig. 17 in the left and right plots, respectively [84]. Also shown in Fig. 17 as light shaded (green) band is the mass

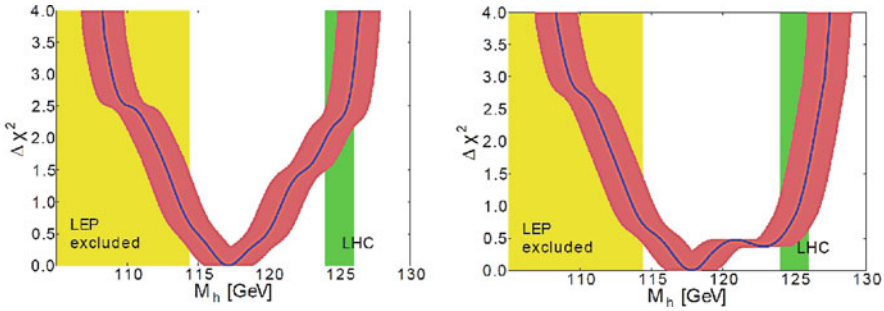

Fig. 17 The $\Delta\chi^2$ functions for M_h in the CMSSM (*left*) and the NUHM1 (*right*) [84], including the theoretical uncertainties (*red bands*). Also shown as light shaded (*green*) band is the mass range corresponding to the newly discovered particle around ∼125 GeV

range corresponding to the newly discovered particle around ∼125 GeV. One can see that the CMSSM is still compatible with $M_h \sim 125$ GeV, while the NUHM1 is in perfect agreement with this light $\mathscr{CP}$-even Higgs boson mass.

Acknowledgements I thank the organizers for their hospitality and for creating a very stimulating environment, in particular during the Whisky tasting.

References

1. P.W. Higgs, Phys. Lett. **12**, 132 (1964); Phys. Rev. Lett. **13**, 508 (1964); Phys. Rev. **145**, 1156 (1966); F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964); G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
2. G. Aad et al., ATLAS Collaboration, Phys. Lett. **B 716**, 1 (2012)
3. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. **B 716**, 30 (2012)
4. Tevatron New Physics Higgs Working Group and CDF and D0 Collaborations (2012), arXiv:1207.0449 [hep-ex]
5. S.L. Glashow, Nucl. Phys. **B 22**, 579 (1961); S. Weinberg, Phys. Rev. Lett. **19**, 19 (1967); A. Salam, in *Proceedings of the 8th Nobel Symposium*, Stockholm, ed. by N. Svartholm (1968)
6. H. Nilles, Phys. Rept. **110**, 1 (1984); H. Haber, G. Kane, Phys. Rept. **117**, 75 (1985); R. Barbieri, Riv. Nuovo Cim. **11**, 1 (1988)
7. S. Heinemeyer, O. Stål, G. Weiglein, Phys. Lett. **B 710**, 201 (2012)
8. S. Weinberg, Phys. Rev. Lett. **37**, 657 (1976); J. Gunion, H. Haber, G. Kane, S. Dawson, *The Higgs Hunter's Guide* (Perseus Publishing, Cambridge, 1990), and references therein
9. P. Fayet, Nucl. Phys. **B 90**, 104 (1975); Phys. Lett. **B 64**, 159 (1976); Phys. Lett. **B 69**, 489 (1977); Phys. Lett. **B 84**, 416 (1979); H.P. Nilles, M. Srednicki, D. Wyler, Phys. Lett. **B 120**, 346 (1983); J.M. Frere, D.R. Jones, S. Raby, Nucl. Phys. **B 222**, 11 (1983); J.P. Derendinger, C.A. Savoy, Nucl. Phys. **B 237**, 307 (1984); J. Ellis, J. Gunion, H. Haber, L. Roszkowski, F. Zwirner, Phys. Rev. **D 39**, 844 (1989); M. Drees, Int. J. Mod. Phys. **A 4**, 3635 (1989); U. Ellwanger, C. Hugonie, A. M. Teixeira, Phys. Rept. **496**, 1 (2010); M. Maniatis, Int. J. Mod. Phys. **A 25**, 3505 (2010)
10. N. Arkani-Hamed, A. Cohen, H. Georgi, Phys. Lett. **B 513**, 232 (2001); N. Arkani-Hamed, A. Cohen, T. Gregoire and J. Wacker, JHEP **0208**, 020 (2002)

11. N. Arkani-Hamed, S. Dimopoulos, G. Dvali, Phys. Lett. **B 429**, 263 (1998) Phys. Lett. **B 436**, 257 (1998); I. Antoniadis, Phys. Lett. **B 246**, 377 (1990); J. Lykken, Phys. Rev. **D 54**, 3693 (1996); L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999)
12. G. Weiglein et al., LHC/ILC Study Group, Phys. Rept. **426**, 47 (2006)
13. A. De Roeck et al., Eur. Phys. J. **C 66**, 525 (2010)
14. LHC2TSP Working Group 1 (EWSB) report, https://indico.cern.ch/contributionDisplay.py?contribId=131&confId=175067
15. N. Cabibbo, L. Maiani, G. Parisi, R. Petronzio, Nucl. Phys. **B 158**, 295 (1979); R. Flores, M. Sher, Phys. Rev. **D 27**, 1679 (1983); M. Lindner, Z. Phys. **C 31**, 295 (1986); M. Sher, Phys. Rept. **179**, 273 (1989); J. Casas, J. Espinosa and M. Quiros, Phys. Lett. **342**, 171 (1995)
16. G. Altarelli and G. Isidori, Phys. Lett. **B 337**, 141 (1994); J. Espinosa and M. Quiros, Phys. Lett. **353**, 257 (1995)
17. T. Hambye and K. Riesselmann, Phys. Rev. **D 55**, 7255 (1997)
18. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections
19. S. Dittmaier et al., LHC Higgs Cross Section Working Group Collaboration (2011), arXiv:1101.0593 [hep-ph]
20. S. Dittmaier et al., LHC Higgs Cross Section Working Group Collaboration (2012), arXiv:1201.3084 [hep-ph]
21. A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, M. Spira, Eur. Phys. J. **C 71**, 1753 (2011)
22. LHC Higgs Cross Section Working Group, A. David et al. (2012), arXiv:1209.0040 [hep-ph]
23. LEP Electroweak Working Group, lepewwg.web.cern.ch/LEPEWWG/Welcome.html
24. Tevatron Electroweak Working Group, tevewwg.fnal.gov
25. The ALEPH, DELPHI, L3 and OPAL Collaborations, the LEP Electroweak Working Group, arXiv:1012.2367 [hep-ex], http://lepewwg.web.cern.ch/LEPEWWG/
26. A. Sirlin, Phys. Rev. **D 22**, 971 (1980); W. Marciano and A. Sirlin, Phys. Rev. **D 22**, 2695 (1980)
27. M. Veltman, Nucl. Phys. **B 123**, 89 (1977)
28. LEP Higgs Working Group, Phys. Lett. **B 565**, 61 (2003)
29. Tevatron Electroweak Working Group and CDF Collaboration and D0 Collaboration, arXiv:1107.5255 [hep-ex]
30. M.W. Grunewald, M.W. Grunewald, J. Phys. Conf. Ser. **110**, 042008 (2008). arXiv:0709.3744 [hep-ph]
31. G. Montagna, O. Nicrosini, F. Piccinini, G. Passarino, Comput. Phys. Commun. **117**, 278 (1999)
32. D. Bardin et al., Comput. Phys. Commun. **133**, 229 (2001); A. Arbuzov et al., Comput. Phys. Commun. **174**, 728 (2006)
33. U. Baur, R. Clare, J. Erler, S. Heinemeyer, D. Wackeroth, G. Weiglein, D. Wood, arXiv:hep-ph/0111314
34. J. Erler, S. Heinemeyer, W. Hollik, G. Weiglein, P. Zerwas, Phys. Lett. **B 486**, 125 (2000); J. Erler, S. Heinemeyer, arXiv:hep-ph/0102083
35. R. Hawkings, K. Mönig, EPJdirect **C8**, 1 (1999)
36. G. Wilson, LC-PHSM-2001-009, www.desy.de/~lcnotes/notes.html
37. A. Hoang et al., Eur. Phys. Jour. **C 3**, 1 (2000); M. Martinez, R. Miquel, Eur. Phys. Jour. **C 27**, 49 (2003)
38. G. Aad et al., The ATLAS Collaboration, arXiv:0901.0512
39. G. Bayatian et al., CMS Collaboration, J. Phys. **G 34**, 995 (2007)
40. S. Heinemeyer et al., arXiv:hep-ph/0511332
41. S. Glashow, S. Weinberg, Phys. Rev. **D 15**, 1958 (1977)
42. B. Petersen, Talk given at HCP2012, http://kds.kek.jp/materialDisplay.py?contribId=46&sessionId=20&materialId=slides&confId=9237; R. Gray, Talk given at HCP2012, http://kds.kek.jp/materialDisplay.py?contribId=48&sessionId=20&materialId=slides&confId=9237
43. A. Djouadi, Phys. Rept. **459**, 1 (2008)
44. J. Gunion, H. Haber, G. Kane, S. Dawson, *The Higgs Hunter's Guide* (Addison-Wesley, Redwood City, 1990)

45. S. Heinemeyer, Int. J. Mod. Phys. **A 21**, 2659 (2006)
46. LEP Higgs Working Group, Eur. Phys. J. **C 47**, 547 (2006)
47. J. Ellis, G. Ridolfi, F. Zwirner, Phys. Lett. **B 257**, 83 (1991); Y. Okada, M. Yamaguchi, T. Yanagida, Prog. Theor. Phys. **85**, 1 (1991); H. Haber, R. Hempfling, Phys. Rev. Lett. **66**, 1815 (1991)
48. A. Brignole, Phys. Lett. **B 281**, 284 (1992)
49. P. Chankowski, S. Pokorski, J. Rosiek, Phys. Lett. **B 286**, 307 (1992); Nucl. Phys. **B 423**, 437 (1994)
50. A. Dabelstein, Nucl. Phys. **B 456**, 25 (1995); Z. Phys. **C 67**, 495 (1995)
51. R. Hempfling, A. Hoang, Phys. Lett. **B 331**, 99 (1994)
52. S. Heinemeyer, W. Hollik, G. Weiglein, Phys. Rev. **D 58**, 091701 (1998); Phys. Lett. **B 440**, 296 (1998)
53. S. Heinemeyer, W. Hollik, G. Weiglein, Eur. Phys. Jour. **C 9**, 343 (1999)
54. R. Zhang, Phys. Lett. **B 447**, 89 (1999); J. Espinosa, R. Zhang, JHEP **0003**, 026 (2000)
55. G. Degrassi, P. Slavich, F. Zwirner, Nucl. Phys. **B 611**, 403 (2001)
56. J. Espinosa and R. Zhang, Nucl. Phys. **B 586**, 3 (2000)
57. A. Brignole, G. Degrassi, P. Slavich, F. Zwirner, Nucl. Phys. **B 631**, 195 (2002)
58. A. Brignole, G. Degrassi, P. Slavich, F. Zwirner, Nucl. Phys. **B 643**, 79 (2002)
59. S. Heinemeyer, W. Hollik, H. Rzehak, G. Weiglein, Eur. Phys. J. **C 39**, 465 (2005)
60. T. Banks, Nucl. Phys. **B 303**, 172 (1988); L. Hall, R. Rattazzi, U. Sarid, Phys. Rev. **D 50**, 7048 (1994); R. Hempfling, Phys. Rev. **D 49**, 6168 (1994); M. Carena, M. Olechowski, S. Pokorski, C. Wagner, Nucl. Phys. **B 426**, 269 (1994)
61. M. Carena, D. Garcia, U. Nierste, C. Wagner, Nucl. Phys. **B 577**, 577 (2000)
62. G. Degrassi, A. Dedes, P. Slavich, Nucl. Phys. **B 672**, 144 (2003)
63. S. Martin, Phys. Rev. **D 65**, 116003 (2002); Phys. Rev. **D 66**, 096001 (2002); Phys. Rev. **D 67**, 095012 (2003); Phys. Rev. **D 68**, 075002 (2003); Phys. Rev. **D 70**, 016005 (2004); Phys. Rev. **D 71**, 016012 (2005); Phys. Rev. **D 71**, 116004 (2005); Phys. Rev. **D 75**, 055005 (2007); S. Martin, D. Robertson, Comput. Phys. Commun. **174**, 133 (2006)
64. R. Harlander, P. Kant, L. Mihaila, M. Steinhauser, Phys. Rev. Lett. **100**, 191602 (2008) [Phys. Rev. Lett. **101**, 039901 (2008)]; JHEP **1008**, 104 (2010)
65. G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich, G. Weiglein, Eur. Phys. J. **C 28**, 133 (2003)
66. S. Heinemeyer, W. Hollik, G. Weiglein, Phys. Rept. **425**, 265 (2006)
67. S. Heinemeyer, W. Hollik, G. Weiglein, Comput. Phys. Commun. **124**, 76 (2000); T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, G. Weiglein, Comput. Phys. Commun. **180**, 1426 (2009), www.feynhiggs.de
68. M. Frank, T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, G. Weiglein, JHEP **0702**, 047 (2007)
69. M. Diaz, H. Haber, Phys. Rev. **D 45**, 4246 (1992)
70. M. Frank, PhD thesis, university of Karlsruhe, 2002
71. M. Carena, S. Heinemeyer, C. Wagner, G. Weiglein, Eur. Phys. J. **C 26**, 601 (2003)
72. T. Hahn, S. Heinemeyer, F. Maltoni, G. Weiglein, S. Willenbrock, arXiv:hep-ph/0607308
73. R. Harlander, S. Liebler, H. Mantler, Computer Physics Communications **184**, 1605 (2013)
74. G. Degrassi and P. Slavich, Nucl. Phys. **B 805**, 267 (2008); G. Degrassi, S. Di Vita and P. Slavich, JHEP **1108**, 128 (2011); Eur. Phys. J. **C 72**, 2032 (2012); R. Harlander, H. Mantler, S. Marzani, K. Ozeren, Eur. Phys. J. **C 66**, 359 (2010); R. Harlander, F. Hofmann, H. Mantler, JHEP **1102**, 055 (2011)
75. M. Carena, S. Heinemeyer, C. Wagner, G. Weiglein, Eur. Phys. J. **C 45**, 797 (2006)
76. CMS Collaboration, CMS-PAS-HIG-12-050, 2012
77. F. Gianotti for the ATLAS collaboration, G. Tonelli for the CMS collaboration, indico.cern.ch/conferenceDisplay.py?confId=164890
78. R. Benbrik, M. Gomez Bock, S. Heinemeyer, O. Stål, G. Weiglein, L. Zeune, Eur. Phys. J. **C 72**, 2171 (2012)
79. P. Bechtle, S. Heinemeyer, O. Stal, T. Stefaniak, G. Weiglein, L. Zeune, Eur. Phys. J. C **73**, 2354 (2013)

80. P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, K.E. Williams, Comput. Phys. Commun. **181**, 138 (2010); Comput. Phys. Commun. **182**, 2605 (2011); P. Bechtle, O. Brein, S. Heinemeyer, O. Stål, T. Stefaniak, G. Weiglein, K. Williams, PoS CHARGED **2012**, 024 (2012)
81. S. Heinemeyer, W. Hollik, D. Stockinger, A. M. Weber, G. Weiglein, JHEP **0608**, 052 (2006)
82. S. Heinemeyer, W. Hollik, G. Weiglein, L. Zeune, JHEP **1312**, 084 (2013). arXiv:1311.1663 [hep-ph]
83. S. AbdusSalam et al., Eur. Phys. J. **C 71**, 1835 (2011)
84. O. Buchmueller et al., Eur. Phys. J. **C 72**, 2243 (2012)

Introduction to Flavour Physics

Yuval Grossman

Abstract This set of lectures covers the very basics of flavour physics and are aimed to be an entry point to the subject. A lot of problems are provided in the hope of making the manuscript a self study guide.

1 Welcome Statement

My plan for these lectures is to introduce you to the very basics of flavour physics. Hopefully, after the lectures you will have enough knowledge, and more importantly, enough curiosity, that you will go on and learn more about the subject.

These are lecture notes and are not meant to be a review. I try to present the basic ideas, hoping to give a clear picture of the physics. Thus, many details are omitted, implicit assumptions are made, and no references are given. Yet the details are important: after you go over the current lecture notes once or twice, I hope you will feel the need for more. This will be the time to turn to the many reviews [1–13] and books [14, 15] on the subject.

I have tried to include many homework problems for the reader to solve, much more than what I gave in the actual lectures. If you would like to learn the material, the provided problems should be a good way to start. They force you to fully understand the issues and apply your knowledge to new situations. The problems are given at the end of each section. The questions can be challenging and may take a lot of time. Do not give up after a few minutes!

Y. Grossman (✉)
Newman Laboratory of Elementary Particle Physics, Department of Physics,
Cornell University, Ithaca, NY 14853, USA
e-mail: yg73@cornell.edu

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_2

2 The Standard Model: A Reminder

I assume that you have basic knowledge of Quantum Field Theory (QFT) and that you are familiar with the Standard Model (SM). Nevertheless, I start with a brief review of the SM, not only to remind you, but also since I like to present things in a way that may be different from the way you got to know the SM.

2.1 The Basics of Model Building

In high energy physics, we ask a very simple question: What are the fundamental laws of Nature? We know that QFT is an adequate tool to describe Nature, at least at energies we have probed so far. Thus the question can be stated in a very compact form as: what is the Lagrangian of nature? The most compact form of the question is

$$\mathscr{L} = ? \tag{1}$$

In order to answer this question we need to provide some axioms or "rules." Our rules are that we "build" the Lagrangian by providing the following three ingredients:

1. The gauge symmetry of the Lagrangian;
2. The representations of fermions and scalars under the symmetry;
3. The pattern of spontaneous symmetry breaking.

The last point practically represented by signs for some parameters.

Once we have specified these ingredients, the next step is to write the most general renormalizable Lagrangian that is invariant under the gauge symmetry and provides the required spontaneous symmetry breaking pattern. This is far from a trivial statement. "Most general" means that all the terms that satisfy the above conditions must be present in the Lagrangian, even the terms that may be phenomenologically problematic. For example, even though we might not want to include a term that induces proton decay, we cannot simply omit it from our model without some symmetry principle that forbids it.

The requirement of renormalizability strongly constrains the form of the Lagrangian and, in fact, limits it to only a finite number of terms. This condition should be thought of an approximation. We think of our model as an effective low energy model and thus write the Lagrangian as a power series in $1/\Lambda$, where Λ is the UV scale.[1] The requirement of renormalizability is therefore equivalent to saying that we truncate the infinite series of operators.

Few remarks are in order about these starting points.

[1]We introduced here some vocabulary. UV refers to short distance or high energy. While we did not use the term IR yet, it worth mentioning that it refers to long distance or low energy.

1. We also impose Poincare invariance. In a way, this can be identified as the gauge symmetry of gravity, and thus can be thought of as part of the first postulate.
2. As we already mentioned, we assume QFT. In particular, quantum mechanics is also an axiom.
3. We do not impose global symmetries. They are accidental, that is, they are there only because we do not allow for non-renormalizable terms.
4. The basic fermion fields are two component Weyl spinors. The basic scalar fields are complex. The vector fields are introduced into the model in order to preserve the gauge symmetry.
5. Any given model has a finite number of parameters. These parameters need to be measured before the model can be tested.

We should elaborate a bit more on the last point. The number of parameters required to define a theory is independent of how they are parametrized. In fact, many not-so-obvious ideas in field theory, such as the renormalization group, become much simpler once one understands that they are reparameterizations of these physical parameters.

In principle, if we consider a theory with k parameters, one would need to make k measurements to establish the theory and then make an extra n ones to probe it. In practice, however, one would conduct all $k + n$ measurements and perform a statistical fit for the k parameters to check for self-consistency.

2.2 An Example: The SM

As an example we consider the SM. It is a nice example, mainly because it describes Nature, and also because the tools we use to construct the SM are also those we use when constructing its possible extensions. The SM is defined as follows:

1. The gauge symmetry is

$$G_{\rm SM} = SU(3)_{\rm C} \times SU(2)_{\rm L} \times U(1)_{\rm Y}. \tag{2}$$

2. There are three fermion generations, each consisting of five representations of $G_{\rm SM}$:

$$Q^I_{Li}(3,2)_{+1/6}, \quad U^I_{Ri}(3,1)_{+2/3}, \quad D^I_{Ri}(3,1)_{-1/3}, \quad L^I_{Li}(1,2)_{-1/2}, \quad E^I_{Ri}(1,1)_{-1}. \tag{3}$$

Our notations mean that, for example, left-handed quarks, Q^I_L, are triplets of $SU(3)_{\rm C}$, doublets of $SU(2)_{\rm L}$ and carry hypercharge $Y = +1/6$. The super-index I denotes gauge interaction eigenstates. The sub-index $i = 1, 2, 3$ is the flavour (or generation) index. There is a single scalar representation,

$$\phi(1,2)_{+1/2}. \tag{4}$$

3. The scalar ϕ assumes a VEV,

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \tag{5}$$

which implies that the gauge group is spontaneously broken,

$$G_{\rm SM} \to SU(3)_{\rm C} \times U(1)_{\rm EM}. \tag{6}$$

This SSB pattern is equivalent to requiring that one parameter in the scalar potential is negative, that is $\mu^2 < 0$, see Eq. (12).

The standard model Lagrangian, $\mathscr{L}_{\rm SM}$, is the most general renormalizable Lagrangian that is consistent with the gauge symmetry (2) and the particle content (3) and (4). It can be divided to three parts:

$$\mathscr{L}_{\rm SM} = \mathscr{L}_{\rm kinetic} + \mathscr{L}_{\rm Higgs} + \mathscr{L}_{\rm Yukawa}. \tag{7}$$

We will learn how to count parameters later, but for now we just mention that $\mathscr{L}_{\rm SM}$ has 18 free parameters[2] that we need to determine experimentally. Now we talk a little about each part of $\mathscr{L}_{\rm SM}$.

For the kinetic terms, in order to maintain gauge invariance, one has to replace the derivative with a covariant derivative:

$$\overline{\psi}\gamma_\mu\partial^\mu\psi \;\to\; \overline{\psi}\gamma_\mu D^\mu\psi, \tag{8}$$

where

$$D^\mu = \partial^\mu + i g_s G_a^\mu L_a + i g W_b^\mu T_b + i g' B^\mu Y. \tag{9}$$

Here G_a^μ are the eight gluon fields, W_b^μ the three weak interaction bosons and B^μ the single hypercharge boson. The L_a's are $SU(3)_{\rm C}$ generators (the 3×3 Gell-Mann matrices $\frac{1}{2}\lambda_a$ for triplets, 0 for singlets), the T_b's are $SU(2)_{\rm L}$ generators (the 2×2 Pauli matrices $\frac{1}{2}\tau_b$ for doublets, 0 for singlets), and the Y's are the $U(1)_{\rm Y}$ charges. For example, for the left-handed quarks, Q_L^I, we have

$$\mathscr{L}_{\rm kinetic}(Q_L) = i\,\overline{Q_{Li}^I}\gamma_\mu\left(\partial^\mu + \frac{i}{2}g_s G_a^\mu\lambda_a + \frac{i}{2}g W_b^\mu\tau_b + \frac{i}{6}g' B^\mu\right)Q_{Li}^I, \tag{10}$$

[2] In fact there is one extra parameter that is related to the vacuum structure of the strong interaction, $\Theta_{\rm QCD}$. Discussing this parameter is far beyond the scope of these lectures.

while for the left-handed leptons, L_L^I, we have

$$\mathscr{L}_{\text{kinetic}}(L_L) = i\overline{L_{Li}^I}\gamma_\mu\left(\partial^\mu + \frac{i}{2}gW_b^\mu\tau_b - \frac{i}{2}g'B^\mu\right)L_{Li}^I. \tag{11}$$

This part of the Lagrangian has three parameters, g, g' and g_s.

The Higgs potential, which describes the scalar self interactions, is given by:

$$-\mathscr{L}_{\text{Higgs}} = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{12}$$

This part of the Lagrangian involves two parameters, λ and μ, or equivalently, the Higgs mass and its VEV. The requirement of vacuum stability tells us that $\lambda > 0$. The pattern of spontaneous symmetry breaking, (5), requires $\mu^2 < 0$. We will not elaborate on this point.

Moving to discuss the Yukawa interactions, we split it into two, the leptonic and baryonic parts. At the renormalizable level the lepton Yukawa interactions are given by

$$-\mathscr{L}_{\text{Yukawa}}^{\text{leptons}} = Y_{ij}^e\overline{L_{Li}^I}\phi E_{Rj}^I + \text{h.c.}. \tag{13}$$

After the Higgs acquires a VEV, these terms lead to charged lepton masses. Note that the SM predicts massless neutrinos. The Lepton Yukawa terms involve three physical parameters, which are usually chosen to be the three charged lepton masses. We will not discuss the lepton sector in these lectures.

The quark Yukawa interactions are given by

$$-\mathscr{L}_{\text{Yukawa}}^{\text{quarks}} = Y_{ij}^d\overline{Q_{Li}^I}\phi D_{Rj}^I + Y_{ij}^u\overline{Q_{Li}^I}\tilde{\phi}U_{Rj}^I + \text{h.c.}, \tag{14}$$

where we introduced here[3] $\tilde{\phi} = i\epsilon_{ij}\phi^*$. This is the part where quarks masses and flavour arises, and we will spend the rest of the lectures on it. For now, just in order to finish the counting, we mention that the Yukawa interactions for the quarks are described by ten physical parameters. They are usually chosen to be the six quark masses and the four parameters of the CKM matrix. We will discuss the CKM matrix at length soon.

2.3 More Symmetries

So far we only mentioned the gauge symmetries that we imposed. Before we go on, let us talk about some other symmetries.

[3]This definition is needed so as to bring the neutral component to be the upper one.

2.3.1 C, P, T and Their Combinations

We start with the discrete symmetries, C, P and T. You may wonder why we discuss these symmetries as we are dealing with flavour. It turns out that in Nature, C, P, and CP violation are closely related to flavour physics. There is no reason for this to be the case, but since it is, we study it simultaneously.

Any local Lorentz invariant QFT conserves CPT, and in particular, this is also the case in the SM. CPT conservation also implies that T violation is equivalent to CP violation. In the SM, C and P are "maximally violated." By that we refer to the fact that both C and P change the chirality of fermion fields. In the SM the left handed and right handed fields have different gauge representations, and thus, independent of the values of the parameters of the model, C and P must be violated in the SM.

The situation with CP is different. The SM can violate CP but it depends on the values of its parameters. It turns out that the parameters of the SM that describe Nature violate CP. The requirement for CP violation is that there is a physical phase in the Lagrangian. In the SM the only place where a complex phase can be physical is in the quark Yukawa interactions. More precisely, in the SM CP is violated if and only if

$$\mathscr{I}m(\det[Y^d Y^{d\dagger}, Y^u Y^{u\dagger}]) \neq 0. \tag{15}$$

An intuitive explanation of why CP violation is related to complex couplings goes as follows. Consider the Yukawa couplings. The Hermiticity of the Lagrangian implies that $\mathscr{L}_{\text{Yukawa}}$ has pairs of terms in the form

$$Y_{ij}\overline{\psi_{Li}}\phi\psi_{Rj} + Y_{ij}^*\overline{\psi_{Rj}}\phi^\dagger\psi_{Li}. \tag{16}$$

A CP transformation exchanges the above two operators

$$\overline{\psi_{Li}}\phi\psi_{Rj} \;\Leftrightarrow\; \overline{\psi_{Rj}}\phi^\dagger\psi_{Li}, \tag{17}$$

but leaves their coefficients, Y_{ij} and Y_{ij}^*, unchanged. This means that $\mathscr{L}_{\text{Yukawa}}$ does not change under CP transformation if $Y_{ij} = Y_{ij}^*$.

In the SM the only source of CP violation are the Yukawa interactions. It is easy to see that the kinetic terms are CP conserving. The SM scalar sector, where there is a single doublet, is also CP conserving. For extended scalar sectors, such as that of a two Higgs doublet model, $\mathscr{L}_{\text{Higgs}}$ can be CP violating.

2.3.2 Global, Accidental and Approximate Symmetries

One might ask how we can have baryon number conservation if our rules above stated that we may not explicitly impose any global symmetries. In general, global symmetries result as outputs of the theory rather than as external constraints. In particular, they come from the structure imposed by renormalizability and gauge

invariance. Global symmetries that appear only because non-renormalizable terms are not considered are called accidental symmetries. These are broken explicitly by non-renormalizable terms, but since these terms are small one can often make use of these symmetries. When people ask why in the SM the proton does not decay, the usual answer is that it is due to baryon number conservation. While this answer is correct, a more precise way to state it is that baryon number violating operators in the SM are non-renormalizable.

The SM has an accidental global symmetry

$$U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau. \tag{18}$$

where $U(1)_B$ is baryon number and the other three $U(1)$s are lepton family lepton numbers. The quarks carry baryon number, while the leptons and the bosons do not. We usually normalize it such that the proton has $B = 1$ and thus each quark carries a third unit of baryon number. As for lepton number, in the SM each family carries its own lepton number, L_e, L_μ and L_τ. Total lepton number is a subgroup of this more general symmetry, that is, the sum of all three family lepton numbers.

In addition to accidental symmetries there are other approximate symmetries which are parametrically small, that is, they become exact global symmetries when a parameter (or a number of parameters) is set to zero. An example is isospin symmetry, which is an approximate global symmetry of the SM and it is broken by the quark masses. Can we consider the quark masses parametrically small? We can only ask if something is small about dimensionless parameters. It turns out that the relevant parameter is the ratio of the quark-mass splitting to the QCD strong coupling scale, e.g.

$$\frac{m_u - m_d}{\Lambda_{\rm QCD}}. \tag{19}$$

Note that isospin symmetry is additionally broken by "the most famous small parameter in the world," $\alpha_{\rm EM}$ since the charge of the u is different from the charge of the d.

There is one more small parameter that you might be familiar with. The Higgs potential obeys a "custodial symmetry." It is broken at one-loop by the Yukawa couplings. While the top Yukawa is not small, $y_t \sim \mathcal{O}(1)$, since the breaking only occurs at loop level the relevant parameter is $y^2/16\pi^2$ which is small.

These accidental and approximate symmetries can be very useful, but we should always remember that they are not as fundamental as the gauge symmetries that we begin with.

2.4 Counting Parameters

Before we go on to study the flavour structure of the SM in detail, we explain how to identify the number of physical parameter in any given model. The Yukawa interactions of Eq. (14) have many parameters but some are not physical. That is, there is a basis where they are identically zero. Of course, it is important to identify the physical parameters in any model in order to probe and check it.

We start with a very simple example. Consider a hydrogen atom in a uniform magnetic field. Before turning on the magnetic field, the hydrogen atom is invariant under spatial rotations, which are described by the $SO(3)$ group. Furthermore, there is an energy eigenvalue degeneracy of the Hamiltonian: states with different angular momenta have the same energy. This degeneracy is a consequence of the symmetry of the system.

When magnetic field is added to the system it makes the direction of the field special. By convention we define this direction as the positive z axis. Let us consider this choice more carefully. A generic uniform magnetic field would be described by three real numbers: the three components of the magnetic field. The magnetic field breaks the $SO(3)$ symmetry of the hydrogen atom system down to an $SO(2)$ symmetry of rotations in the plane perpendicular to the magnetic field. The one generator of the $SO(2)$ symmetry is the only valid symmetry generator now, since the remaining two $SO(3)$ generators in the orthogonal directions are broken. These broken symmetry generators allow us to rotate the system such that the magnetic field points in the z direction:

$$O_{xz}O_{yz}(B_x, B_y, B_z) = (0, 0, B_{z'}), \tag{20}$$

where O_{xz} and O_{yz} are rotations in the xz and yz planes respectively. The two broken generators were used to rotate away two unphysical parameters, leaving us with one physical parameter, the magnitude of the magnetic field. That is, when turning on the magnetic field, all measurable quantities in the system depend only on one new parameter, rather than the naïve three.

The results described above are more generally applicable. In particular, they are useful in studying the flavour physics of quantum field theories. Consider a gauge theory with some matter content. This theory always has kinetic and gauge terms, which have a certain global symmetry, G_f, on their own. Upon adding a potential that respects the gauge symmetries, the global symmetry may be broken down to a smaller symmetry group. When breaking the global symmetry there is some freedom to rotate away unphysical parameters, as when a magnetic field is added to the hydrogen atom system.

In order to analyze this process we define a few quantities. The added potential has coefficients that can be described by N_{general} parameters in a general basis. The global symmetry of the entire model, H_f, has fewer generators than G_f and we call the difference in the number of generators N_{broken}. Finally, the quantity that we would ultimately like to determine is the number of parameters affecting physical measurements, N_{phys}. These numbers are related by

$$N_{\text{phys}} = N_{\text{general}} - N_{\text{broken}}. \tag{21}$$

Furthermore, the rule in (21) applies separately to both real parameters (masses and mixing angles) and phases.

The rule given by (21) can be applied to the flavour sector of the SM. Consider the quark sector of the model. The kinetic term has a global symmetry

$$G_f = U(3)_Q \times U(3)_U \times U(3)_D. \tag{22}$$

A $U(3)$ has 9 generators (3 real and 6 imaginary), so the total number of generators of G_f is 27. The Yukawa interactions defined in (14), Y^F $(F = u, d)$, are 3×3 complex matrices, which contain a total of 36 parameters (18 real parameters and 18 phases) in a general basis. These parameters also break G_f down to baryon number

$$U(3)_Q \times U(3)_U \times U(3)_D \to U(1)_B. \tag{23}$$

While $U(3)^3$ has 27 generators, $U(1)_B$ has only one and thus $N_{\text{broken}} = 26$. This broken symmetry allows us to rotate away a large number of the parameters by moving to a more convenient basis. Using (21), the number of physical parameters should be given by

$$N_{\text{phys}} = 36 - 26 = 10. \tag{24}$$

These parameters can be split into real parameters and phases. The three unitary matrices generating the symmetry of the kinetic and gauge terms have a total of 9 real parameters and 18 phases. The symmetry is broken down to a symmetry with only one phase generator. Thus,

$$N^{(r)}_{\text{phys}} = 18 - 9 = 9, \qquad N^{(i)}_{\text{phys}} = 18 - 17 = 1. \tag{25}$$

We interpret this result by saying that of the 9 real parameters, 6 are the fermion masses and three are the CKM matrix mixing angles. The one phase is the CP-violating phase of the CKM mixing matrix.

In your homework you will count the number of parameters for different models.

2.5 Homework

Question 1 (Global symmetries). We talked about the fact that global symmetries are accidental in the SM, that is, they are broken once non-renormalizable terms are included. Write a lowest dimension term that break each of the global symmetries of the SM. For the lepton numbers it will be of dimension 5, while for the baryon number dimension 6.

Question 2 (Lepton counting). Count the number of physical flavour parameters in the lepton sector and show that there are just three real parameters that we may identify as the masses of the three charged leptons.

Question 3 (Extra generations counting). Count the number of physical flavour parameters in the quark sector of an extended SM with n generations. Show that such a model has $n(n+3)/2$ real parameters and $(n-1)(n-2)/2$ complex phases. Identify the real parameters as masses and mixing angles and determine how many mixing angles there are.

2.6 The CKM Matrix

We are now equipped with the basic concepts that are necessary in order to study the Yukawa interactions. A further tool we will need is that of basis rotations. There are two important bases. One where the masses are diagonal, called the mass basis, and the other where the $W^\pm$ interactions are diagonal, called the interaction basis. The fact that these two bases are not the same results in flavour changing interactions. The CKM matrix is the matrix that rotates between these two bases.

Since most measurements are done in the mass basis, we write the interactions in that basis. Upon the replacement $\mathscr{R}e(\phi^0) \to (v + H^0)/\sqrt{2}$, see Eq. (5), we decompose the $SU(2)_{\rm L}$ quark doublets into their components:

$$Q^I_{Li} = \begin{pmatrix} U^I_{Li} \\ D^I_{Li} \end{pmatrix}, \tag{26}$$

and then the Yukawa interactions, Eq. (14), give rise to mass terms:

$$-\mathscr{L}^q_M = (M_d)_{ij}\,\overline{D^I_{Li}}\,D^I_{Rj} + (M_u)_{ij}\,\overline{U^I_{Li}}\,U^I_{Rj} + \text{h.c.}, \qquad M_q = \frac{v}{\sqrt{2}}Y^q\,. \tag{27}$$

The mass basis corresponds, by definition, to diagonal mass matrices. We can always find unitary matrices V_{qL} and V_{qR} such that

$$V_{qL}M_qV^\dagger_{qR} = M^{\rm diag}_q, \qquad q = u, d, \tag{28}$$

with $M^{\rm diag}_q$ diagonal and real. The quark mass eigenstates are then identified as

$$q_{Li} = (V_{qL})_{ij}q^I_{Lj}, \qquad q_{Ri} = (V_{qR})_{ij}q^I_{Rj}, \qquad q = u, d. \tag{29}$$

The charged current interactions for quarks are the interactions of the $W^\pm_\mu$, which in the interaction basis are described by (10). They have a more complicated form in the mass basis:

$$-\mathscr{L}^{q}_{W^{\pm}} = \frac{g}{\sqrt{2}}\overline{u_{Li}}\gamma^{\mu}(V_{uL}V^{\dagger}_{dL})_{ij}\,d_{Lj}\,W^{+}_{\mu} + \text{h.c.}. \tag{30}$$

The unitary 3×3 matrix,

$$V = V_{uL}V^{\dagger}_{dL}, \qquad (VV^{\dagger} = \mathbf{1}), \tag{31}$$

is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix for quarks. As a result of the fact that V is not diagonal, the $W^{\pm}$ gauge bosons couple to mass eigenstates quarks of different generations. Within the SM, this is the only source of flavour changing interactions.

The form of the CKM matrix is not unique. We usually fix part of this freedom by ordering the up quarks and the down quarks by their masses: $(u_1, u_2, u_3) \to (u, c, t)$ and $(d_1, d_2, d_3) \to (d, s, b)$. The elements of V are therefore written as follows:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{32}$$

Then, we can use many parametrizations. The standard parametrization, used by the Particle Data Group (PDG) [16], is given by

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{33}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three $\sin\theta_{ij}$ are the three real mixing parameters while δ is called the Kobayashi-Maskawa phase. Another parametrization is the Wolfenstein parametrization where the four mixing parameters are (λ, A, ρ, η) where η represents the CP violating phase. The Wolfenstein parametrization is an expansion in the small parameter, $\lambda = |V_{us}| \approx 0.22$. To $O(\lambda^3)$ the parametrization is given by

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}. \tag{34}$$

We will talk in detail about how to measure the CKM parameters. For now let us mention that the Wolfenstein parametrization is a good approximation to the actual numerical values. That is, the CKM matrix is very close to a unit matrix with off diagonal terms that are small. The order of magnitude of each element can be deduced from the power of λ in the Wolfenstein parametrization.

Various parametrizations differ in the way that the freedom of phase rotation is used to leave a single phase in V. One can define, however, a CP violating quantity in V that is independent of the parametrization. This quantity, the Jarlskog invariant, $J_{\rm CKM}$, is defined through

$$\mathscr{I}m(V_{ij}V_{kl}V_{il}^*V_{kj}^*) = J_{\rm CKM}\sum_{m,n=1}^{3}\epsilon_{ikm}\epsilon_{jln}, \qquad (i,j,k,l=1,2,3). \tag{35}$$

In terms of the explicit parametrizations given above, we have

$$J_{\rm CKM} = c_{12}c_{23}c_{13}^2 s_{12}s_{23}s_{13}\sin\delta \approx \lambda^6 A^2\eta. \tag{36}$$

The condition (15) can be translated to the language of the flavour parameters in the mass basis. Then we see that a necessary and sufficient condition for CP violation in the quark sector of the SM:

$$\Delta m_{tc}^2\Delta m_{tu}^2\Delta m_{cu}^2\Delta m_{bs}^2\Delta m_{bd}^2\Delta m_{sd}^2 J_{\rm CKM} \neq 0. \tag{37}$$

where we define $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$. Equation (37) puts the following requirements on the SM in order that it violates CP:

1. Within each quark sector, there should be no mass degeneracy;
2. None of the three mixing angles should be zero or $\pi/2$;
3. The phase should be neither 0 nor π.

A very useful concept is that of the unitarity triangles. The unitarity of the CKM matrix leads to various relations among the matrix elements, for example,

$$\sum_i V_{id}V_{is}^* = 0. \tag{38}$$

There are six such relations and they require the sum of three complex quantities to vanish. Therefore, they can be geometrically represented in the complex plane as a triangle and are called "unitarity triangles." It is a feature of the CKM matrix that all unitarity triangles have equal areas. Moreover, the area of each unitarity triangle equals $|J_{\rm CKM}|/2$ while the sign of $J_{\rm CKM}$ gives the direction of the complex vectors around the triangles.

One of these triangles, the one corresponding to the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \tag{39}$$

has sides of roughly the same lengths and is relatively easy to probe. For these reasons, Eq. (39) is referred to as "*the* unitarity triangle". We further define the rescaled unitarity triangle. It is derived from Eq. (39) by choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real and dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. The rescaled unitarity triangle is similar to the unitarity triangle. Two vertices of the rescaled

unitarity triangle are fixed at (0,0) and (1,0). The coordinates of the remaining vertex correspond to the Wolfenstein parameters (ρ, η). The lengths of the two complex sides are

$$R_u \equiv \left|\frac{V_{ud}V_{ub}}{V_{cd}V_{cb}}\right| = \sqrt{\rho^2+\eta^2}, \qquad R_t \equiv \left|\frac{V_{td}V_{tb}}{V_{cd}V_{cb}}\right| = \sqrt{(1-\rho)^2+\eta^2}. \tag{40}$$

The three angles of the unitarity triangle are defined as follows:

$$\alpha \equiv \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \qquad \beta \equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \qquad \gamma \equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{41}$$

They are physical quantities and can be independently measured, as we will discuss later. Another commonly used notation is $\phi_1 = \beta$, $\phi_2 = \alpha$, and $\phi_3 = \gamma$. Note that in the standard parametrization $\gamma = \delta_{\rm KM}$.

2.7 FCNCs

So far we have talked about flavour-changing charged currents that are mediated by the $W^\pm$ bosons. In the SM, this is the only source of flavour changing interaction and, in particular, of generation changing interaction. There is no fundamental reason why there cannot be flavour Changing Neutral Currents (FCNCs). After all, two interactions of flavour changing charged current result in a neutral current interaction. Yet, experimentally we see that FCNCs processes are highly suppressed.

This is a good place to pause and open your PDG.[4] Look, for example, at the rate for the neutral current decay, $K_L \to \mu^+\mu^-$, and compare it to that of the charged current decay, $K^+ \to \mu^+\nu$. You see that the K_L decay rate is much smaller. Browse the PDG a bit more and see that the same pattern is found in D and B decays.

The fact that the data show that FCNCs are highly suppressed implies that any model that aims to describe Nature must have a mechanism to suppress FCNCs. The way the SM deals with the data is to make sure there are no tree level FCNCs. In the SM, all neutral current interactions are flavour conserving at the tree level and FCNCs are mediated only at the loop level and are therefore suppressed (we discuss the exact amount of suppression below).

Before proceeding let us make a short remark. We often distinguish between non-diagonal couplings, diagonal couplings and universal couplings. Universal couplings are diagonal couplings with the same strength. An important point to

[4] It goes without saying that every student in high energy physics must have the PDG [16]. If, for some reason you do not have one, order it now. It is free and has a lot of important stuff. Until you get it, you can use the online version at pdg.lbl.gov.

note is that universal couplings are diagonal in any basis. Non-universal diagonal couplings, in general, become non-diagonal after a basis rotation.

There are four types of neutral bosons in the SM that could mediate tree-level neutral currents. They are the gluons, photon, Higgs and Z bosons. We study each of them in turn, explain what is required in order to make their couplings diagonal in the mass basis, and how this requirement is fulfilled in the SM.

We start with the massless gauge bosons: the gluons and photon. For these, tree-level couplings are always diagonal, independently of the details of the theory. The reason is that these gauge bosons correspond to exact gauge symmetries. Thus, their couplings to the fermions arise from the kinetic terms. When the kinetic terms are canonical, the couplings of the gauge bosons are universal and, in particular, flavour conserving. In other words, gauge symmetry plays a dual role: it guarantees that the gauge bosons are massless and that their couplings are flavour universal.

Next consider the Higgs interactions. The reason that the Higgs couplings are diagonal in the SM is that the Higgs couplings to fermions are aligned with the mass matrix. This alignment is a consequence of the fact that both the Higgs couplings to fermions and the fermion mass matrices are proportional to the same Yukawa couplings. To see that this is the case we consider the Yukawa interactions (14). After inserting $\mathscr{R}e(\phi^0) \to (v+h)/\sqrt{2}$ and keeping both the fermion masses and Higgs fermion interaction terms we get

$$\begin{aligned} -\mathscr{L}_{\text{Yukawa}}^{\text{quarks}} &= Y_{ij}^d \overline{Q_{Li}^I} \phi D_{Rj}^I + Y_{ij}^u \overline{Q_{Li}^I} \tilde{\phi} U_{Rj}^I \\ &= \frac{Y_{ij}^d}{\sqrt{2}} \left(\overline{D_{Li}^I} D_{Rj}^I \right) (v+h) + \frac{Y_{ij}^u}{\sqrt{2}} \left(\overline{U_{Li}^I} U_{Rj}^I \right) (v+h). \end{aligned} \tag{42}$$

Diagonalizing the mass matrix, we get the interaction in the physical basis

$$M_d \left(\overline{D_{Li}} D_{Rj} \right) (v+h) + M_u \left(\overline{U_{Li}} U_{Rj} \right) (v+h). \tag{43}$$

Clearly, since everything is proportional to $(v+h)$, the interaction is diagonalized together with the mass matrix.

This special feature of the Higgs interaction is tightly related to the facts that the SM has only one Higgs field and that the only source for fermion masses is the Higgs VEV. In models where there are additional sources for the masses, that is, bare mass terms or more Higgs fields, diagonalization of the mass matrix does not simultaneously diagonalize the Higgs interactions. In general, there are Higgs mediated FCNCs in such models. In your homework you will work out an example.

Last, we discuss Z-mediated FCNCs. The coupling for the Z to fermions is proportional to $T_3 - q \sin^2 \theta_W$ and in the interaction basis the Z couplings to quarks are given by

$$-\mathscr{L}_Z = \frac{g}{\cos\theta_W}\left[\overline{u^I_{Li}}\gamma^\mu\left(\frac{1}{2}-\frac{2}{3}\sin^2\theta_W\right)u^I_{Li} + \overline{u^I_{Ri}}\gamma^\mu\left(-\frac{2}{3}\sin^2\theta_W\right)u^I_{Ri}\right.$$
$$\left.+\overline{d^I_{Li}}\gamma^\mu\left(-\frac{1}{2}+\frac{1}{3}\sin^2\theta_W\right)d^I_{Li}+\overline{d^I_{Ri}}\gamma^\mu\left(\frac{1}{3}\sin^2\theta_W\right)d^I_{Ri}\right]Z_\mu + \text{h.c.}. \quad (44)$$

In order to demonstrate the fact that there are no Z-mediated FCNCs at tree level, let us concentrate only on the left-handed up-type quarks. Rotating to the mass eigenstates we find

$$-\mathscr{L}_Z = \frac{g}{\cos\theta_W}\left[\overline{u_{Li}}\ (V_{uL})_{ik}\ \gamma^\mu\left(\frac{1}{2}-\frac{2}{3}\sin^2\theta_W\right)\left(V^\dagger_{uL}\right)_{kj}\ u_{Lj}\right],$$
$$= \frac{g}{\cos\theta_W}\left[\overline{u_{Li}}\,\gamma^\mu\left(\frac{1}{2}-\frac{2}{3}\sin^2\theta_W\right)u_{Li}\right] \quad (45)$$

where in the last step we used

$$V_{uL}V^\dagger_{uL} = 1. \quad (46)$$

We see that the interaction is universal and diagonal in flavour. It is easy to verify that this holds for the other types of quarks. Note the difference between the neutral and the charged currents cases. In the neutral current case we insert $V_{uL}V^\dagger_{uL} = 1$. This is in contrast to the charged current interactions where the insertion is $V_{uL}V^\dagger_{dL}$, which in general is not equal to the identity matrix.

The fact that there are no FCNCs in Z-exchange is due to some specific properties of the SM. That is, we could have Z-mediated FCNCs in simple modifications of the SM. The general the condition for the absence of tree level FCNCs is as follows. Fields can mix if they belong to the same representation under all the unbroken generators. That is, they must have the same spin, electric charge and $\mathrm{SU}(3)_C$ representation. If these fields also belong to the same representation under the broken generators their couplings to the massive gauge boson is universal. If, however, they belong to different representations under the broken generators, their couplings in the interaction basis are diagonal but non-universal. These couplings become non-diagonal after rotation to the mass basis.

In the SM, the requirement mentioned above for the absence of Z-exchange FCNCs is satisfied. That is, all the fields that belong to the same representation under the unbroken generators also belong to the same representation under the broken generators. For example, all left handed quarks with electric charge $2/3$ also have the same hypercharge $(1/6)$ and they are all an up component of a double of $SU(2)_L$ and thus have $T_3 = 1/2$. This does not have to be the case. After all, $Q = T_3 + Y$, so there are many ways to get quarks with the same electric charge. In your homework, you will work out the details of a model with non-standard representations and see how it exhibits Z-exchange FCNCs.

2.8 Homework

Question 4 (Exotic light quarks). We consider a model with the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ spontaneously broken by a single Higgs doublet into $SU(3)_C \times U(1)_{EM}$. The quark sector, however, differs from the standard model one as it consists of three quark flavours, that is, we do not have the c, b and t quarks. The quark representations are non-standard. Of the left handed quarks, $Q_L = (u_L, d_L)$ form a doublet of $SU(2)_L$ while s_L is a singlet. All the right handed quarks are singlets. All color representations and electric charges are the same as in the standard model.

1. Write down (a) the gauge interactions of the quarks with the charged W bosons (before SSB); (b) the Yukawa interactions (before SSB); (c) the bare mass terms (before SSB); (d) the mass terms after SSB.
2. Show that there are four physical flavour parameters in this model. How many are real and how many imaginary? Is there CP violation in this model? Separate the parameters into masses, mixing angles and phases.
3. Write down the gauge interactions of the quarks with the Z boson in both the interaction basis and the mass basis. (You do not have to rewrite terms that do not change when you rotate to the mass basis. Write only the terms that are modified by the rotation to the mass basis.) Are there, in general, tree level Z exchange FCNCs? (You can assume CP conservation from now on.)
4. Are there photon and gluon mediated FCNCs? Support your answer by an argument based on symmetries.
5. Are there Higgs exchange FCNCs?
6. Repeat the question with a somewhat different model, where the only modification is that two of the right handed quarks, $Q_R = (u_R, d_R)$, form a doublet of $SU(2)_L$. Note that there is one relation between the real parameters that makes the parameter counting a bit tricky.

Question 5 (Two Higgs doublet model). Consider the two Higgs doublet model (2HDM) extension of the SM. In this model, we add a Higgs doublet to the SM fields. Namely, instead of the one Higgs field of the SM we now have two, denoted by ϕ_1 and ϕ_2. For simplicity you can work with two generations when the third generation is not explicitly needed.

1. Write down (in a matrix notation) the most general Yukawa potential of the quarks.
2. Carry out the diagonalization procedure for such a model. Show that the Z couplings are still flavour diagonal.
3. In general, however, there are FCNCs in this model, mediated by the Higgs bosons. To show that, write the Higgs fields as $\mathrm{Re}(\phi_i) = v_i + h_i$ where $i = 1, 2$ and $v_i \neq 0$ is the VEV of ϕ_i, and define $\tan\beta = v_2/v_1$. Then, write down the

Higgs–fermion interaction terms in the mass basis. Assuming that there is no mixing between the Higgs fields, you should find a non-diagonal Higgs fermion interaction terms.

3 Probing the CKM

Now that we have an idea about flavour in general and in the SM in particular, we are ready to compare the SM predictions with data. While we use the SM as an example, the tools and ideas are applicable to a large number of theories.

The basic idea is as follows. In order to check a model we first have to determine its parameters and then we can probe it. When considering the flavour sector of the SM, this implies that we first have to measure the parameters of the CKM matrix and then check the model. We can think about the first four independent measurements as determining the CKM parameters and the fifth measurements on are then used to probe the SM. In practice, however, the SM is checked by looking for consistency among many measurements. Any inconsistency is a signal of new physics.[5]

There is one major issue that we need to think about: how precisely can the predictions of the theory be tested? Our ability to test any theory is limited by uncertainties. There are two kinds of uncertainties: experimental and theoretical. There are many sources of both kinds, and a lot of research has gone into trying to overcome them in order to be able to better probe the SM and its extensions.

We do not elaborate on experimental details. We just make one general point. Since our goal is to probe the small elements of the CKM, we have to measure very small branching ratios, typically down to $O(10^{-6})$. To do that we need a lot of statistics and a superb understanding of the detectors and the backgrounds.

As for theory errors, there is basically one player here: QCD, or, using its mighty name, "the strong interaction." Yes, it is strong, and yes, it is a problem for us. Basically, we can only deal with weakly coupled forces. The use of perturbation theory is so fundamental to our way of doing physics. It is very hard to deal with phenomena that cannot be described with perturbation theory.

In practice the problem is that our theory is given in terms of quarks, but measurements are done with hadrons. It is far from trivial bridge this gap. In particular, it becomes hard when we are looking for high precision. There are basically two ways to overcome the problem of QCD. One way is to find observables for which the needed hadronic input can be measured or eliminated. The other way is to use approximate symmetries of QCD, in particular, isospin, $SU(3)_F$ and heavy quark symmetries. Below we only mention how these are used without getting into much detail.

[5]The term "new physics" refers to any model that extends the SM. We are eager to find indications for new physics and determine what that new physics is. At the end of the lectures we expand on this point.

3.1 Measuring the CKM Parameters

When we attempt to determine the CKM parameters we talk about two classifications. One classification is related to what we are trying to extract:

1. Measure magnitudes of CKM elements or, equivalently, sides of the unitarity triangle;
2. Measure phases of CKM elements or, equivalently, angles of the unitarity triangle;
3. Measure combinations of magnitudes and phases.

The other classification is based on the physics, in particular, we classify measurements based on the type of amplitudes that are involved:

1. Tree-level amplitudes. Such measurements are also referred to as "direct measurements;"
2. Loop amplitudes. Such measurements are also referred to as "indirect measurements;"
3. Cases where both tree-level and loop amplitude are involved.

There is no fundamental difference between direct and indirect measurement. We make the distinction since direct measurements are expected to be valid in most extensions of the SM. Such models have a special flavour structure that suppresses flavour-changing couplings and have a very small effect on processes that are large in the SM, which are generally tree-level processes. In contrast new physics can have a large effect on processes that are very small in the SM, mainly loop processes. Thus, we refer to loop amplitude measurements as indirect measurements.

3.2 Direct Measurements

In order to determine the magnitudes of CKM elements, a number of sophisticated theoretical and experimental techniques are needed, the complete discussion of which is beyond the scope of these lectures. Instead, we give one example, the determination of $|V_{cb}|$ and, hopefully, you will find the time to read about direct determinations of other CKM parameters in one of the reviews such as the PDG or Ref. [9].

The basic idea in direct determination of CKM elements is to use the fact that the amplitudes of semi-leptonic tree-level decays are proportional to one CKM element. In the case of $b \to c\ell\bar{\nu}$ it is proportional to V_{cb}. (While the diagram is not given here, it is a good opportunity to verify that you can draw it and see how the dependence on the CKM element enters.) Experimentally, it turns out that $|V_{ub}| \ll |V_{cb}|$. Therefore we can neglect the $b \to u\ell\bar{\nu}$ decays and use the full semileptonic B decays data set to measure $|V_{cb}|$ without the need to know the hadronic final state.

The way to overcome the problem of QCD is to use heavy quarks symmetry (HQS). We do not discuss the use of HQS in detail here. We just mention that the small expansion parameter is Λ_{QCD}/m_b. The CKM element $|V_{cb}|$ can be extracted from inclusive and exclusive semileptonic B decays.

In the inclusive case, the problem is that the calculation is done using the b and c quarks. In particular, the largest uncertainty is due to the fact that at the quark level the decay rate scales like m_b^5. The definition of the b quark mass, as well as the measurements of it, is complicated: How can we define the mass of a particle that is never free? All we can define and measure very precisely is the B meson mass.[6] Using an operator product expansion (OPE) together with the heavy quark effective theory, we can expand in the small parameter Λ_{QCD}/m_b and get a reasonable estimate of $|V_{cb}|$. The point to emphasize is that this is a controllable expansion, that is, we know that

$$\Gamma(b \to c\ell\bar{\nu}) = \Gamma(B \to X_c\ell\bar{\nu})\left(1 + \sum_n a_n\right), \tag{47}$$

such that a_n is suppressed by $(\Lambda_{QCD}/m_B)^n$. In principle we can calculate all the a_n and get a very precise prediction. It is helpful that $a_1 = 0$. The calculation of a_n has been done for $n = 2$ and $n = 3$.

The exclusive approach overcomes the problem of the need for the b quark mass by looking at specific hadronic decays, in particular, $B \to D\ell\bar{\nu}$ and $B \to D^*\ell\bar{\nu}$. Here the problem is that these decays cannot be calculated in terms of quarks: it has to be done in terms of hadrons. This is where using "form factors" is useful as we now explain. The way to think about the problem is that a b field creates a free b quark or annihilates a free anti-b quark. Yet, inside the meson the b is not free. Thus the operator that we care about, $\bar{c}\gamma_\mu b$, is not directly related to annihilating the b quark inside the meson. The mismatch is parametrized by form factors. The form factors are functions of the momentum transfer. In general, we need some model to calculate these form factors, as they are related to the strong interaction. In the B case, we can use HQS, which tells us that in the limit $m_B \to \infty$ all the form factors are universal, and are given by the (unknown) Isgur-Wise function. The fact that we know something about the form factors makes the theoretical errors rather small, below the 5 % level.

Similar ideas are used when probing other CKM elements. For example, in β-decay the $d \to ue\bar{\nu}$ decay amplitude is proportional to V_{ud}. Here the way to overcome QCD is by using isospin where the expansion parameter is m_q/Λ_{QCD} with $q = u, d$. Another example is K-decay, $s \to ue\bar{\nu} \propto V_{us}$. In that case, in addition to isospin, flavour SU(3) is used where we assume that the strange quark is light. In some cases, this is a good approximation, but not as good as isospin.

[6]There is an easy way to remember the mass of the B meson that is based on the fact that it is easier to remember two things than one. It is rather amusing to note that the number of feet there are in one mile and the mass of the B meson in MeV are, in fact, the same, 5,280.

Direct measurements have been used to measure the magnitude of seven out of the nine CKM matrix components. The two exceptions are $|V_{ts}|$ and $|V_{td}|$. The reason is that the world sample of top decays is very small, and moreover, it is very hard to determine the flavour of the light quark in top decay. These two elements are best probed using loop processes, as we discuss next.

3.3 Indirect Measurements

The CKM dependence of decay amplitudes involved in direct measurements of the CKM elements is simple. The amplitudes are tree level with one internal W propagator. In the case of semileptonic decays, the amplitude is directly proportional to one CKM matrix element.

The situation with loop decays is different. Usually we concentrate on FCNC[7] processes at the one loop level. Since the loop contains an internal W propagator, we gain sensitivity to CKM elements. The sensitivity is always to a combination of CKM elements. Moreover, there are several amplitudes with different internal quarks in the loop. These amplitudes come with different combinations of CKM elements. The total amplitude is the sum of these diagrams, and thus it has a non-trivial dependence on a combination of CKM elements.

As an example consider one of the most interesting loop induced decay, $b \to s\gamma$. There are several amplitudes for this decay, which you should try to draw. Basically, they are all one loop diagrams. Note that we have to sum over all possible internal quarks. Each set of diagrams with a given internal up-type quarks, u_i, is proportional to $V_{ib}V_{is}^*$. It can further depend on the mass of the internal quark. Thus, we can write the total amplitude as

$$A(b \to s\gamma) \propto \sum_{i=u,c,t} f(m_i) V_{ib} V_{is}^*. \tag{48}$$

While the expression in (48) looks rather abstract, we can gain a lot of insight into the structure of the amplitude by recalling that the CKM matrix is unitary. Using

$$\sum_{i=u,c,t} V_{ib} V_{is}^* = 0, \tag{49}$$

we learn that the contribution of the m_i independent term in f vanishes. An explicit calculation shows that $f(m_i)$ grows with m_i and that the leading term in the expansion in powers of m_i/m_W scales as m_i^2.

[7] In the first lecture we proved that in the SM there are no tree-level FCNCs. Why do we talk about FCNCs here? I hope the answer is clear.

The fact that in loop decays the amplitude is proportional to m_i^2/m_W^2 is called the GIM mechanism. Historically, it was the first theoretical motivation of the charm quark. Before the charm was discovered, it was a puzzle that the decay $K_L \to \mu^+\mu^-$ was not observed. The GIM mechanism provided an answer. The fact that the CKM is unitary implies that this process is a one loop process and there is an extra suppression of order m_c^2/m_W^2 to the amplitude. Thus, the rate is tiny and very hard to observe.

The GIM mechanism is also important in understanding the finiteness of loop amplitudes. Any one-loop amplitude corresponding to a decay where the tree-level amplitude vanishes must be finite. Technically, this can be seen by noticing that if it were divergence, a counter term at tree-level would be needed, but that cannot be the case if the tree-level amplitude vanishes. The amplitude for $b \to s\gamma$ it is naively log divergent (it is important that you verify this by power counting). Yet, it is only the m_i independent term that diverges. The GIM mechanism is here to save us as it guarantees that this term is zero. The m_i-dependent result is finite, as it should be.

One more important point about the GIM mechanism is the fact that the amplitude is proportional to the mass squared of the internal quark. This implies that the heavier the internal quark is, the more sensitive is the total amplitude to its couplings. In B decays, the heaviest internal quark is the top quark. This is the reason that $b \to s\gamma$ is sensitive to $|V_{ts}V_{tb}|$. This is a welcome feature since, as we mentioned before, these elements are hard to probe directly.

In one-loop decays of kaons, there is a "competition" between the internal top and charm quarks. The top is heavier, but the CKM couplings of the charm are larger. Numerically, the charm is the winner, but not by a large margin. Check for yourself.

As for charm decay, since the tree-level decay amplitudes are large, and since there is no heavy internal quark, the loop decay amplitudes are highly suppressed. So far the experimental bounds on various loop-mediated charm decays are far above the SM predictions. As an exercise, try to determine which internal quark dominates the one-loop charm decay.

3.4 Homework

Question 6 (Direct CKM measurements from *D* decays). The ratio of CKM elements

$$r \equiv \frac{|V_{cd}|}{|V_{cs}|} \tag{50}$$

can be estimated assuming SU(3) flavour symmetry. The idea is that in the SU(3) limit the pion and the kaon have the same mass and the same hadronic matrix elements.

1. Construct a ratio of semileptonic D decays that can be used to measure the ratio r.
2. We usually expect SU(3) breaking effects to be of the order $m_s/\Lambda_{QCD} \sim 20\,\%$. Compare the observable you constructed to the actual measurement and estimate the SU(3) breaking effect.

Question 7 (The GIM mechanism: $b \to s\gamma$ decay).

1. Explain why $b \to s\gamma$ is a loop decay and draw the one-loop diagrams in the SM.
2. Naively, these diagrams diverge. Show this.
3. Once we add all the diagrams and make use of the CKM unitarity, we get a finite result. Show that the UV divergences cancel (that is, assume that all the masses are the same and show that the answer is zero).
4. We now add a vector-like pair of down-type quarks to the SM which we denote by b'

$$b'_R(3,1)_{-1/3}, \qquad b'_L(3,1)_{-1/3}. \tag{51}$$

   Show that in that model Eq. (49) is not valid anymore, that is,

$$\sum_{i=u,c,t} V_{ib}V_{is}^* \neq 0, \tag{52}$$

   and that we have a Z exchange tree level FCNCs in the down sector. (The name "vector-like" refers to the case where the left and right handed fields have the same representation under all gauge groups. This is in contrast to a chiral pair where they have different representations. All the SM fermions are chiral.)
5. As we argued, one cannot have $b \to s\gamma$ at tree level in any model. Thus, in the model with the vector-like quarks, the one-loop diagrams must also be finite. Yet, in the SM we used Eq. (49) to argue that the amplitude is finite, but now it is not valid. Show that the amplitude is finite also in this case. (Hint: When you have an infinite result that should be finite the reason is usually that there are more diagrams that you forgot.)

4 Meson Mixing

Another interesting FCNC process is neutral meson mixing. Since it is a FCNC process, it cannot be mediated at tree level in the SM, and thus it is related to the "indirect measurements" class of CKM measurements. Yet, the importance of meson mixing and oscillation goes far beyond CKM measurements and we study it in some detail.

4.1 Formalism

There are four neutral mesons that can mix: K, D, B, and B_s.[8] We first study the general formalism and then the interesting issues in each of the systems. The formalism is that of a two body open system. That is, the system involves the meson states P^0 and $\overline{P}^0$, and all the states they can decay to. Before the meson decays the state can be a coherent superposition of the two meson states. Once the decay happens, coherence is practically lost. This allows us to describe the decays using a non-Hermitian Hamiltonian, like we do for an open system.

We consider a general meson denoted by P. At $t = 0$ it is in an initial state

$$|\psi(0)\rangle = a(0)|P^0\rangle + b(0)|\overline{P}^0\rangle \ , \tag{53}$$

where we are interested in computing the values of $a(t)$ and $b(t)$. Under our assumptions all the evolution is determined by a 2×2 effective Hamiltonian $\mathscr{H}$ that is not Hermitian. Any complex matrix, such as $\mathscr{H}$, can be written in terms of Hermitian matrices M and Γ as

$$\mathscr{H} = M - \frac{i}{2}\,\Gamma \ . \tag{54}$$

M and Γ are associated with $(P^0, \overline{P}^0) \leftrightarrow (P^0, \overline{P}^0)$ transitions via off-shell (dispersive) and on-shell (absorptive) intermediate states, respectively. The diagonal elements of M and Γ are associated with the flavour-conserving transitions $P^0 \to P^0$ and $\overline{P}^0 \to \overline{P}^0$, while the off-diagonal elements are associated with flavour-changing transitions $P^0 \to \overline{P}^0$ and $\overline{P}^0 \to P^0$.

If $\mathscr{H}$ is not diagonal, the meson states P^0 and $\overline{P}^0$ are not mass eigenstates, and thus do not have well-defined masses and decay widths. It is only the eigenvectors of $\mathscr{H}$ that have well-defined masses and widths. We denote[9] the light and heavy eigenstates as P_L and P_H with $m_H > m_L$. Note that since $\mathscr{H}$ is not Hermitian, the eigenvectors do not need to be orthogonal to each other. Due to CPT, $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. Then when we solve the eigenvalue problem for $\mathscr{H}$ we find that the eigenstates are given by

$$|P_{L,H}\rangle = p|P^0\rangle \pm q|\overline{P}^0\rangle, \tag{55}$$

[8] You may be wondering why there are only four meson mixing systems. If you do not wonder and do not know the answer, then you should wonder. We will answer this question shortly.

[9] Another possible choice, which is standard for K mesons, is to define the mass eigenstates according to their lifetimes: K_S for the short-lived and K_L for the long-lived state. The K_L is experimentally found to be the heavier state.

with the normalization $|p|^2 + |q|^2 = 1$ and

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - (i/2)\Gamma_{12}^*}{M_{12} - (i/2)\Gamma_{12}} \,. \tag{56}$$

If CP is a symmetry of $\mathscr{H}$ then M_{12} and Γ_{12} are relatively real ($\mathscr{I}m(\Gamma_{12}/M_{12}) = 0$), and therefore

$$\left|\frac{q}{p}\right| = 1 \,, \tag{57}$$

where the phase of q/p is unphysical. In this case the mass eigenstates are orthogonal

$$\langle P_H | P_L \rangle = |p|^2 - |q|^2 = 0 \,. \tag{58}$$

The real and imaginary parts of the eigenvalues of $\mathscr{H}$ corresponding to $|P_{L,H}\rangle$ represent their masses and decay-widths, respectively. The mass difference Δm and the width difference $\Delta\Gamma$ are defined as follows:

$$\Delta m \equiv M_H - M_L, \qquad \Delta\Gamma \equiv \Gamma_H - \Gamma_L. \tag{59}$$

Note that here Δm is positive by definition, while the sign of $\Delta\Gamma$ is to be determined experimentally. (Alternatively, one can use the states defined by their lifetimes to have $\Delta\Gamma \equiv \Gamma_S - \Gamma_L$ positive by definition.) The average mass and width are given by

$$m \equiv \frac{M_H + M_L}{2}, \qquad \Gamma \equiv \frac{\Gamma_H + \Gamma_L}{2}. \tag{60}$$

It is useful to define dimensionless ratios x and y:

$$x \equiv \frac{\Delta m}{\Gamma}, \qquad y \equiv \frac{\Delta\Gamma}{2\Gamma}. \tag{61}$$

We also define

$$\theta = \arg(M_{12}\Gamma_{12}^*). \tag{62}$$

Solving the eigenvalue equation gives

$$(\Delta m)^2 - \frac{1}{4}(\Delta\Gamma)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2), \qquad \Delta m \Delta\Gamma = 4\mathscr{R}e(M_{12}\Gamma_{12}^*). \tag{63}$$

In the limit of CP conservation, Eq. (63) is simplified to

$$\Delta m = 2|M_{12}|, \qquad |\Delta\Gamma| = 2|\Gamma_{12}|. \tag{64}$$

4.2 Time Evolution

We move on to study the time evolution of a neutral meson. For simplicity, we assume CP conservation. Later on, when we study CP violation, we will relax this assumption, and study the system more generally. Many important points, however, can be understood in the simplified case when CP is conserved and so we use it here.

In the CP limit $|q| = |p| = 1/\sqrt{2}$ and we can choose the relative phase between p and q to be zero. In that case the transformation from the flavour to the mass basis, (55), is simplified to

$$|P_{L,H}\rangle = \frac{1}{\sqrt{2}}\left(|P^0\rangle \pm |\overline{P}^0\rangle\right). \tag{65}$$

We denote the state of an initially pure $|P^0\rangle$ after an time t as $|P^0(t)\rangle$ (and similarly for or $|\overline{P}^0\rangle$). We obtain

$$|P^0(t)\rangle = \cos\left(\frac{\Delta E\, t}{2}\right)|P^0\rangle + i\,\sin\left(\frac{\Delta E\, t}{2}\right)|\overline{P}^0\rangle\,, \tag{66}$$

and similarly for $|\overline{P}^0(t)\rangle$. Since flavour is not conserved, the probability to measure a specific flavour, that is P or $\overline{P}$, oscillates in time, and it is given by

$$\begin{aligned}\mathscr{P}(P \to P)[t] &= \left|\langle P^0(t)|P^0\rangle\right|^2 = \frac{1+\cos(\Delta E t)}{2},\\ \mathscr{P}(P \to \overline{P})[t] &= \left|\langle P^0(t)|\overline{P}^0\rangle\right|^2 = \frac{1-\cos(\Delta E t)}{2},\end{aligned} \tag{67}$$

where $\mathscr{P}$ denotes probability.

A few remarks are in order:

- In the meson rest frame, $\Delta E = \Delta m$ and $t = \tau$, the proper time.
- We learn that we have flavour oscillation with frequency Δm. This is the parameter that eventually gives us the sensitivity to the weak interaction and to flavour.
- We learn that by measuring the oscillation frequency we can determine the mass splitting between the two mass eigenstates. One way this can be done is by measuring the flavour of the meson at both production and decay. It is not trivial to measure the flavour at both ends, and we do not describe it in detail here, but you are encouraged to think and learn about how it can be done.

4.3 Time Scales

Next, we spend some time understanding the different time scales that are involved in meson mixing. One scale is the oscillation period. As can be seen from Eq. (67), the oscillation time scale is given by Δm.[10]

Before we talk about the other time scales we have to understand how the flavour is measured, or as we usually say it, tagged. By "flavour is tagged" we refer to the decay as a flavour vs anti-flavour, for example b vs $\bar{b}$. Of course, in principle, we can tag the flavour at any time. In practice, however, the measurement is done for us by Nature. That is, the flavour is tagged when the meson decays. In fact, it is done only when the meson decays in a flavour-specific way. Other decays that are common to both P and $\overline{P}$ do not measure the flavour. Such decays are also very useful as we will discuss later. Semi-leptonic decays are very good flavour tags:

$$b \to c\mu^- \bar{\nu}, \qquad \bar{b} \to \bar{c}\mu^+ \nu. \tag{68}$$

The charge of the lepton tells us the flavour: a μ^+ tells us that we "measured" a b flavour, while a μ^- indicates a $\bar{b}$. Of course, before the meson decays it could be in a superposition of a b and a $\bar{b}$. The decay acts as a quantum measurement. In the case of a semileptonic decay, it acts as a measurement of flavour vs anti-flavour.

Aside from the oscillation time, one other relevant time scale is the time when the flavour measurements is done. Since the flavour is tagged when the meson decays, the relevant time scale is the decay width, Γ. We can then use the dimensionless quantity, $x \equiv \Delta m/\Gamma$, defined in (61), to understand the relevance of these two time scales. There are three relevant regimes:

1. $x \ll 1$. We denote this case as "slow oscillation". In that case the meson has no time to oscillate, and thus to good approximation flavour is conserved. In practice, this implies that $\cos(\Delta m t) \approx 1$ and using it in Eq. (67) we see that $\mathscr{P}(P \to P) \approx 1$ and $\mathscr{P}(P \to \overline{P}) \to 0$. In this case, an upper bound on the mass difference is likely to be established before an actual measurement. This case is relevant for the D system.
2. $x \gg 1$. We denote this case as "fast oscillation". In this case the meson oscillates many times before decaying, and thus the oscillating term practically averages out to zero.[11] In practice in this case $\mathscr{P}(P \to P) \approx \mathscr{P}(P \to \overline{P}) \approx 1/2$ and a lower bound on Δm can be established before a measurement can be done. This case is relevant for the B_s system.

[10] What we refer to here is, of course, $1/\Delta m$. Yet, at this stage of our life as physicists, we know how to match dimensions, and thus I interchange between time and energy freely, counting on you to understand what I am referring to.

[11] This is the case we are very familiar with when we talk about decays into mass eigenstates. There is never a decay into a mass eigenstate. Only when the oscillations are very fast and the oscillatory term in the decay rate averages out, the result seems like the decay is into a mass eigenstate.

3. $x \sim 1$. In this case the oscillation and decay times are roughly the same. That is, the system has time to oscillate and the oscillation are not averaged out. In a way, this is the most interesting case since then it is relatively easy to measure Δm. Amazingly, this case is relevant to the K and B systems. We emphasize that the physics that leads to Γ and Δm are unrelated, so there is no reason to expect $x \sim 1$. Yet, Nature is kind enough to produce $x \sim 1$ in two out of the four neutral meson systems.

It is amusing to point out that oscillations give us sensitivity to mass differences of the order of the width, which are much smaller than the mass itself. In fact, we have been able to measure mass differences that are 14 orders of magnitude smaller than the corresponding masses. It is due to the quantum mechanical nature of the oscillation that such high precision can be achieved.

In some cases there is one more time scale: $\Delta\Gamma$. In such cases, we have one more relevant dimensionless parameter $y \equiv \Delta\Gamma/(2\Gamma)$. Note that y is bounded, $-1 \leq y \leq 1$. (This is in contrast to x which is bounded from below, $x > 0$.) Thus, we can talk about several cases depending on the values of y and x.

1. $|y| \ll 1$ and $y \ll x$. In this case the width difference is irrelevant. This is the case for the B system.
2. $y \sim x$. In this case the width difference is as important as the oscillation. This is the case in the D system where $y \ll 1$ and for the K system with $y \sim 1$.
3. $|y| \sim 1$ and $y \ll x$. In this case the oscillation averages out and the width different shows up as a difference in the lifetime of the two mass eigenstates. This case may be relevant to the B_s system, where we expect $y \sim 0.1$.

There are few other limits (like $y \gg x$) that are not realized in the four meson systems. Yet, they might be realized in some other systems yet to be discovered.

To conclude this subsection we summarize the experimental data on meson mixing

$$\begin{aligned} x_K &\sim 1, & y_K &\sim 1, \\ x_D &\sim 10^{-2}, & y_D &\sim 10^{-2}, \\ x_d &\sim 1, & y_d &\lesssim 10^{-2}, \\ x_s &\sim 10, & y_s &\lesssim 10^{-1}, \end{aligned} \tag{69}$$

where the latter two refer to the B_d and B_s systems, respectively. Note that y_d and y_s have not been measured and all we have are upper bounds.

4.4 Calculation of the Mixing Parameters

We will now explain how the calculation of the mixing parameters is done. We will only make some brief comments on $\Delta\Gamma$ and then spend some time on the calculation of Δm. As we have done a few times, we will do the calculation in

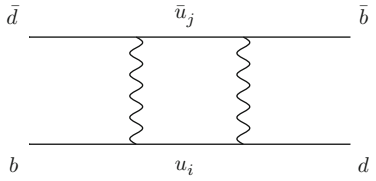

Fig. 1 A box diagram that generates an operator that can lead to a $B \leftrightarrow \overline{B}$ transition

the SM, keeping in mind that the tools we develop can be used in a large class of models.

In order to calculate the mass and width differences, we need to know the effective Hamiltonian, $\mathscr{H}$, defined in Eq. (54). For the diagonal terms, no calculations are needed. CPT implies $M_{11} = M_{22}$ and to an excellent approximation it is just the mass of the meson. Similarly, $\Gamma_{11} = \Gamma_{22}$ is the average width of the meson. What we need to calculate is the off diagonal terms, that is M_{12} and Γ_{12}.

We start by discussing M_{12}. For the sake of simplicity we consider the B meson as a concrete example. The first point to note is that M_{12} is basically the transition amplitude between a B and a $\overline{B}$ at zero momentum transfer. In terms of states with the conventional normalization we have

$$M_{12} = \frac{1}{2m_B}\langle B|\mathscr{O}|\overline{B}\rangle. \tag{70}$$

We emphasize that we should not square the amplitude. We square amplitudes to get transition probabilities and decay rates, which is not the case here.

The operator that appears in (70) is one that can create a B and annihilate a $\overline{B}$. Recalling that a B meson is made of a $\bar{b}$ and d quark (and $\overline{B}$ from b and $\bar{d}$), we learn that in terms of quarks it must be of the form[12]

$$\mathscr{O} \sim (\bar{b}\,d)(\bar{b}\,d). \tag{71}$$

Since the operator in (71) is a FCNC operator, in the SM it cannot be generated at tree level and must be generated at one loop. The one loop diagram that generates it is called "a box diagram", because it looks like a square. It is given in Fig. 1. The calculation of the box diagram is straightforward and we end up with

$$M_{12} \propto \frac{g^4}{m_W^2}\langle B|(\bar{b}_L\gamma_\mu d_L)(\bar{b}_L\gamma^\mu d_L)|\overline{B}\rangle \sum_{i,j} V_{id}^* V_{ib} V_{jd}^* V_{jb} F(x_i, x_j), \tag{72}$$

[12] We do not explicitly write the Dirac structure. Anything that does not vanish is possible.

such that

$$x_i \equiv \frac{m_i^2}{m_W^2}, \qquad i = u, c, t, \tag{73}$$

and the function F is known, but we do not write it here. Several comments are in order:

1. The box diagram is second order in the weak interaction, that is, it is proportional to g^4.
2. The fact that the CKM is unitary (in other words, the GIM mechanism) makes the m_i independent term vanish and to a good approximation $\sum_{i,j} F(x_i, x_j) \to F(x_t, x_t)$. We then say that it is the top quark that dominates the loop.
3. The final ingredient we need is the hadronic matrix element,

$$\langle B|(\bar{b}_L \gamma_\mu d_L)(\bar{b}_L \gamma^\mu d_L)|\overline{B}\rangle.$$

   The problem is that the operator creates a free b and $\bar{d}$ quark and annihilates a free $\bar{b}$ and a d. This is not the same as creating a $\overline{B}$ meson and annihilating a B meson. Here, lattice QCD helps and by now a good estimate of the matrix element is available.
4. Similar calculations can be done for the other mesons. Due to the GIM mechanism for the K meson the function F gives an extra m_c^2/m_W^2 suppression.
5. Last we mention the calculation of Γ_{12}. An estimate of it can be made by looking at the on-shell part of the box diagram. Yet, once an intermediate particle goes on shell, QCD becomes important, and the theoretical uncertainty in the calculation of Γ_{12} is larger than that of M_{12}.

Putting all the pieces together we see how the measurement of the mass difference is sensitive to some combination of CKM elements. Using the fact that the amplitude is proportional to the heaviest internal quark we get from (72) and (64)

$$\Delta m_B \propto |V_{tb} V_{td}|^2, \tag{74}$$

where the proportionality constant is known with an uncertainty at the 10 % level.

4.5 Homework

Question 8 (The four mesons). It is now time to come back and ask why there are only four mesons that we care about when discussing oscillations. In particular, why we do not talk about oscillation for the following systems

1. $B^+ - B^-$ oscillation
2. $K - K^*$ oscillation

3. $T - \overline{T}$ oscillation (a T is a meson made out of a t and a $\bar{u}$ quarks.)
4. $K^* - \overline{K}^*$ oscillation

Hint: The last three cases all have to do with time scales. In principle there are oscillations in these systems, but they are irrelevant.

Question 9 (Kaons). Here we study some properties of the kaon system. We did not talk about it at all. You have to go back and recall (or learn) how kaons decay, and combine that with what we discussed in the lecture.

1. Explain why $y_K \approx 1$.
2. In a hypothetical world where we could change the mass of the kaon without changing any other masses, how would the value of y_K change if we made m_K smaller or larger.

Question 10 (Mixing beyond the SM). Consider a model without a top quark, in which the first two generations are as in the SM, while the left–handed bottom (b_L) and the right–handed bottom (b_R) are $SU(2)$ singlets.

1. Draw a tree-level diagram that contributes to $B - \bar{B}$ mixing in this model.
2. Is there a tree-level diagram that contributes to $K - \bar{K}$ mixing?
3. Is there a tree-level diagram that contributes to $D - \bar{D}$ mixing?

5 CP Violation

As we already mentioned, it turns out that in Nature CP violation is closely related to flavour. In the SM, this is manifested by the fact that the source of CP violation is the phase of the CKM matrix. Thus, we will spend some time learning about CP violation in the SM and beyond.

5.1 How to Observe CP Violation?

CP is the symmetry that relates particles with their anti-particles. Thus, if CP is conserved, we must have

$$\Gamma(A \to B) = \Gamma(\bar{A} \to \bar{B}), \tag{75}$$

such that A and B represent any possible initial and final states. From this we conclude that one way to find CP violation is to look for processes where

$$\Gamma(A \to B) \neq \Gamma(\bar{A} \to \bar{B}). \tag{76}$$

This, however, is not easy. The reason is that even when CP is not conserved, Eq. (75) can hold to a very high accuracy in many cases. So far there are only very few cases where (75) does not hold to a measurable level. The reason that it is not easy to observe CP violation is that there are several conditions that have to be fulfilled. CP violation can arise only in interference between two decay amplitudes. These amplitudes must carry different weak and strong phases (we explain below what these phases are). Also, CPT implies that the total width of a particle and its anti-particle are the same. Thus, any CP violation in one channel must be compensated by CP violation with an opposite sign in another channel. Finally, it happens that in the SM, which describes Nature very well, CP violation comes only when we have three generations, and thus any CP violating observable must involve all the three generations. Due to the particular hierarchical structure of the CKM matrix, all CP violating observables are proportional to very small CKM elements.

In order to show this we start by defining weak and strong phases. Consider, for example, the $B \to f$ decay amplitude A_f, and the CP conjugate process, $\bar{B} \to \overline{f}$, with decay amplitude $\overline{A}_{\overline{f}}$. There are two types of phases that may appear in these decay amplitudes. Complex parameters in any Lagrangian term that contributes to the amplitude will appear in complex conjugate form in the CP-conjugate amplitude. Thus, their phases appear in A_f and $\overline{A}_{\overline{f}}$ with opposite signs and these phases are CP odd. In the SM, these phases occur only in the couplings of the $W^{\pm}$ bosons and hence CP odd phases are often called "weak phases."

A second type of phases can appear in decay amplitudes even when the Lagrangian is real. They are from possible contributions of intermediate on-shell states in the decay process. These phases are the same in A_f and $\overline{A}_{\overline{f}}$ and are therefore CP even. One type of such phases is easy to calculate. It comes from the trivial time evolution, $\exp(iEt)$. More complicated cases are where there is rescattering due to the strong interactions. For this reason these phases are called "strong phases."

There is one more kind of phase in addition to the weak and strong phases discussed so far. These are the spurious phases that arise due to an arbitrary choice of phase convention, and do not originate from any dynamics. For simplicity, we set these unphysical phases to zero from now on.

It is useful to write each contribution a_i to A_f in three parts: its magnitude $|a_i|$, its weak phase ϕ_i, and its strong phase δ_i. If, for example, there are two such contributions, $A_f = a_1 + a_2$, we have

$$\begin{aligned} A_f &= |a_1| e^{i(\delta_1+\phi_1)} + |a_2| e^{i(\delta_2+\phi_2)}, \\ \overline{A}_{\overline{f}} &= |a_1| e^{i(\delta_1-\phi_1)} + |a_2| e^{i(\delta_2-\phi_2)}. \end{aligned} \tag{77}$$

Similarly, for neutral meson decays, it is useful to write

$$M_{12} = |M_{12}| e^{i\phi_M}, \qquad \Gamma_{12} = |\Gamma_{12}| e^{i\phi_\Gamma}. \tag{78}$$

Each of the phases appearing in Eqs. (77) and (78) is convention dependent, but combinations such as $\delta_1 - \delta_2$, $\phi_1 - \phi_2$, and $\phi_M - \phi_\Gamma$ are physical. Now we can see why in order to observe CP violation we need two different amplitudes with different weak and strong phases. This is easy to show and I leave it as homework. A few remarks are in order:

1. The basic idea in CP violation research is to find processes where we can measure CP violation. That is, we look for processes with two decay amplitudes that are roughly of the same size with different weak and strong phases.
2. In some cases, we can get around QCD. In such cases, we get sensitivity to the phases of the unitarity triangle (or, equivalently, of the CKM matrix). These cases are the most interesting ones.
3. Some observables are sensitive to CP phases without measuring CP violation. That is like saying that we can determine the angles of a triangle just by knowing the lengths of its sides.
4. While we talk only about CP violation in meson oscillations and decays, there are other types of CP-violating observables. In particular, triple products and electric dipole moments (EDMs) of elementary particles encode CP violation. They are not directly related to flavour, and are not covered here.
5. So far CP violation has been observed only in meson decays, particularly, in K_L, B_d and $B^\pm$ decays. In the following, we concentrate on the formalism relevant to these systems.

5.2 The Three Types of CP Violation

When we consider CP violation in meson decays there are two types of amplitudes: mixing and decay. Thus, there must be three ways to observe CP violation, depending on which type of amplitudes interfere. Indeed, this is the case. We first introduce the three classes and then discuss each of them in some length.

1. CP violation in decay, also called direct CP violation. This is the case when the interference is between two decay amplitudes. The necessary strong phase is due to rescattering.
2. CP violation in mixing, also called indirect CP violation. In this case the absorptive and dispersive mixing amplitudes interfere. The strong phase is due to the time evolution of the oscillation.
3. CP violation in interference between mixing and decay. As the name suggests, here the interference is between the decay and the oscillation amplitudes. The dominant effect is due to the dispersive mixing amplitude (the one that gives the mass difference) and a leading decay amplitude. Also here the strong phase is due to the time evolution of the oscillation.

In all of the above cases the weak phase comes from the Lagrangian. In the SM these weak phases are related to the CKM phase. In many cases, the weak phase is one of the angles of the unitary triangle.

5.3 CP Violation in Decay

Let us first discuss CP violation in decay. This is the case when

$$|A(P \to f)| \neq |A(\overline{P} \to \overline{f})|. \tag{79}$$

The way to measure this type of CP violation is as follows. We define

$$a_{CP} \equiv \frac{\Gamma(\bar{B} \to \bar{f}) - \Gamma(B \to f)}{\Gamma(\bar{B} \to \bar{f}) + \Gamma(B \to f)} = \frac{|\bar{A}/A|^2 - 1}{|\bar{A}/A|^2 + 1}. \tag{80}$$

Using (77) with ϕ as the weak phase difference and δ as the strong phase difference, we write

$$A(P \to f) = A\,(1 + r\exp[i(\phi+\delta)])\,, \qquad A(\overline{P} \to \overline{f}) = A\,(1 + r\exp[i(-\phi+\delta)])\,, \tag{81}$$

with $r \leq 1$. We get

$$a_{CP} = 2r\sin\phi\sin\delta. \tag{82}$$

This result shows explicitly that we need two decay amplitudes, that is, $r \neq 0$, with different weak phases, $\phi \neq 0, \pi$ and different strong phases $\delta \neq 0, \pi$. A few remarks are in order:

1. In order to have a large effect we need each of the three factors in (82) to be large.
2. CP violation in decay can occur in both charged and neutral mesons. One complication in the case of a neutral meson is that it is not always possible to tell the flavour of the decaying meson, that is, whether it is P or $\overline{P}$. This can be a problem or a virtue.
3. In general the strong phase is not calculable since it is related to QCD. This may not be a problem if all we are after is to demonstrate CP violation. In other cases the phase can be independently measured, eliminating this particular source of theoretical error.

5.3.1 $B \to K\pi$

Our first example of CP violation in decay is $B^0 \to K^+\pi^-$. At the quark level the decay is mediated by $b \to s\bar{u}u$ transition. There are two dominant decay amplitudes,

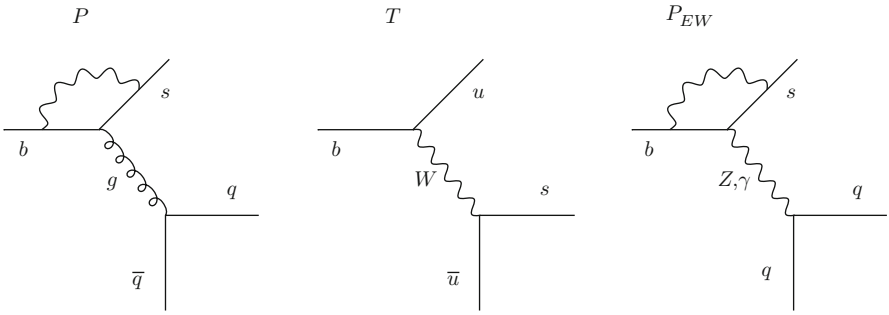

Fig. 2 The $B \to K\pi$ amplitudes. The dominant one is the strong penguin amplitude (P), and the sub-dominant ones are the tree amplitude (T) and the electroweak penguin amplitude (P_{EW})

tree level and one loop penguin diagrams.[13] Two penguin diagrams and the tree level diagram are plotted in Fig. 2.

Naively, tree diagrams are expected to dominate. Yet, this is not the case here. The reason is that the tree diagram is highly CKM suppressed. It turns out that this suppression is stronger than the loop suppression such that $r = |P/T| \sim 0.3$, where we use P and T to denote the penguin and tree amplitudes, respectively. In terms of weak phases, the tree amplitude carries the phase of $V_{ub}V_{us}^*$. The dominant internal quark in the penguin diagram is the top quark and thus to first approximation the phase of the penguin diagram is the phase of $V_{tb}V_{ts}^*$, and to first approximation $\phi = \alpha$. As for the strong phase, we cannot calculate it, and there is no reason for it to vanish since the two amplitudes have a different structure. Experimentally, CP violation in $B \to K\pi$ decays has been established. It was the first observation of CP violation in decay.

We remark that $B \to K\pi$ decays have much more to offer. There are four different such decays, and they are all related by isospin, and thus many predictions can be made. Moreover, the decay rates are relatively large and the measurements have been performed. The full details are beyond the scope of these lectures, but you are encouraged to go and study them.

5.3.2 $B \to DK$

Our second example is $B \to DK$ decay. This decay involves only tree level diagrams, and is sensitive to the phase between the $b \to c\bar{u}s$ and $b \to u\bar{c}s$ decay amplitudes, which is γ. The situation here is involved as the D further decays and

[13] This is the first time we introduce the name penguin. It is just a name, and it refers to a one-loop amplitude of the form $f_1 \to f_2 B$ where B is a neutral boson that can be on–shell or off–shell. If the boson is a gluon we may call it QCD penguin. When it is a photon or a Z boson it is called an electroweak penguin.

what is measured is $B \to f_D K$, where f_D is a final state of a D or $\overline{D}$ decay. This "complication" turns out to be very important. It allows us to construct theoretically very "clean" observables. In fact, $B \to DK$ decays are arguably the cleanest measurement of a CP violation phase in terms of theoretical uncertainties.

The reason for this theoretical cleanliness is that all the necessary hadronic quantities can be extracted experimentally. We consider decays of the type

$$B \to D(\overline{D})\, K(X) \to f_D\, K(X), \tag{83}$$

where f_D is a final state that can be accessible from both D and $\overline{D}$ and X represents possible extra particles in the final state. The crucial point is that in the intermediate state the flavour is not measured. That is, the state is in general a coherent superposition of D and $\overline{D}$. On the other hand, this state is on-shell, so the $B \to D$ and $D \to f_D$ amplitudes factorize. Thus, we have quantum coherence and factorization at the same time. The coherence makes it possible to have interference and thus sensitivity to CP violating phases. Factorization is important since then we can separate the decay chain into stages such that each stage can be determined experimentally. The combination is then very powerful, we have a way to probe CP violation without the need to calculate any decay amplitude.

To see the power of the method, consider using $B \to DKX$ decays with n different X states, and $D \to f_D$ with k different f_D states, one can perform $n \times k$ different measurements. Because the B and D decay amplitude factorize, these $n \times k$ measurements depend on $n + k$ hadronic decay amplitudes. For large enough n and k, there is a sufficient number of measurements to determine all hadronic parameters, as well as the weak phase we are after. Since all hadronic matrix elements can be measured, the theoretical uncertainties are much below the sensitivity of any foreseeable future experiment.

5.4 CP Violation that Involves Mixing

We move on to study CP violation that involves mixing. This kind of CP violation is the one that was first discovered in the kaon system in the 1960s, and in the B system more recently. They are the ones that shape our understanding of the picture of CP violation in the SM, and thus, they deserve some discussion.

We start by re-deriving the oscillation formalism in the more general case where CP violation is included. Then we will be able to construct some CP violating observables and see how they are related to the phases of the unitarity triangle. For simplicity we concentrate on the B system. We allow the decay to be into an arbitrary state, that is, a state that can come from any mixture of B and $\overline{B}$. Consider a final state f such that

$$A_f \equiv A(B \to f), \qquad \bar{A}_f \equiv A(\overline{B} \to f). \tag{84}$$

We further define

$$\lambda_f \equiv \frac{q}{p}\frac{\overline{A}_f}{A_f} \,. \tag{85}$$

We consider the general time evolution of a P^0 and $\overline{P}^0$ mesons. It is given by

$$\begin{aligned} |P^0(t)\rangle &= g_+(t)\,|P^0\rangle - (q/p)\; g_-(t)|\overline{P}^0\rangle, \\ |\overline{P}^0(t)\rangle &= g_+(t)\,|\overline{P}^0\rangle - (p/q)\; g_-(t)|P^0\rangle \,, \end{aligned} \tag{86}$$

where we work in the B rest frame and

$$g_\pm(t) \equiv \frac{1}{2}\left(e^{-im_H t-\frac{1}{2}\Gamma_H t} \pm e^{-im_L t-\frac{1}{2}\Gamma_L t}\right). \tag{87}$$

We define $\tau \equiv \Gamma t$ and then the decay rates are

$$\begin{aligned} \Gamma(B \to f)[t] = |A_f|^2 e^{-\tau} \;&\Big\{(\cosh y\tau + \cos x\tau) + |\lambda_f|^2(\cosh y\tau - \cos x\tau) \\ &-2\mathrm{Re}\left[\lambda_f(\sinh y\tau + i\,\sin x\tau)\right]\Big\}, \\ \Gamma(\bar{B} \to f)[t] = |\bar{A}_f|^2 e^{-\tau} \;&\Big\{(\cosh y\tau + \cos x\tau) + |\lambda_f|^{-2}(\cosh y\tau - \cos x\tau) \\ &-2\mathrm{Re}\left[\lambda_f^{-1}(\sinh y\tau + i\,\sin x\tau)\right]\Big\}, \end{aligned} \tag{88}$$

where $\Gamma(B \to f)[t]$ ($\Gamma(\overline{B} \to f)[t]$) is the probability for an initially pure B ($\overline{B}$) meson to decay at time t to a final state f.

Terms proportional to $|A_f|^2$ or $|\overline{A}_f|^2$ are associated with decays that occur without any net oscillation, while terms proportional to $|\lambda_f|^2$ or $|\lambda_f|^{-2}$ are associated with decays following a net oscillation. The $\sinh(y\tau)$ and $\sin(x\tau)$ terms in Eqs. (88) are associated with the interference between these two cases. Note that, in multi-body decays, amplitudes are functions of phase-space variables. The amount of interference is in general a function of the kinematics, and can be strongly influenced by resonant substructure. Equations (88) are much simplified in the case where $|q/p| = 1$ and $|A_f/\overline{A}_f| = 1$. In that case λ_f is a pure phase, i.e. $|\lambda_f| = 1$.

We define the CP observable of the asymmetry of neutral meson decays into a final state f by

$$\mathscr{A}_f(t) \equiv \frac{\Gamma[\overline{B}(t) \to \overline{f}] - \Gamma[B(t) \to f]}{\Gamma[\overline{B}(t) \to \overline{f}] + \Gamma[B(t) \to f]}. \tag{89}$$

We restrict ourself to the case where f is a CP eigenstates, and then $f = \overline{f}$. Also consider the case where there is no CP violation in the decay, and then the

decay amplitudes fulfill $|\overline{A}_f| = |A_f|$. We further consider the case where $\Delta\Gamma = 0$ and $|q/p| = 1$, as expected to a good approximation for B system. In that case the interference between decays with and without mixing is the only source of the asymmetry and we get

$$\mathscr{A}_f(t) = \mathscr{I}m(\lambda_f)\sin(x\tau) = \sin\left[\arg(\lambda_f)\right]\sin(\Delta m t). \tag{90}$$

where in the last step we used $|\lambda_f| = 1$. We see that once we know Δm, and if the above conditions are satisfied, we have a clean measurement of the phase of λ_f. This phase is directly related to an angle in the unitarity triangle, as we discuss shortly.

It is instructive to describe the effect of CP violation in decays of the mass eigenstates. For cases where the width difference is negligible, this is usually not very useful. It is not easy to generate, for example, a B_H mass eigenstate. When the width difference is large, as in the kaon system, this representation can be very useful since we do know how to generate K_L states. Assuming $|q/p| = 1$, the decay into CP eigenstates is given by

$$\Gamma(P_L \to f_{CP}) = 2|A_f|^2 e^{-\Gamma_L t}\cos^2\theta, \qquad \Gamma(P_H \to f_{CP}) = 2|A_f|^2 e^{-\Gamma_H t}\sin^2\theta, \tag{91}$$

where

$$\theta \equiv \frac{\arg\lambda_f}{2}. \tag{92}$$

In this case, CP violation, that is $\theta \neq 0$, is manifested by the fact that two non-degenerate states can decay to the same CP eigenstate final state.

We also consider decays into a pure flavour state. In that case $\lambda_f = 0$, and we can isolate the effect of CP violation in mixing. CP violation in mixing is defined by

$$|q/p| \neq 1\ . \tag{93}$$

This is the only source of CP violation in charged-current semileptonic neutral meson decays $P, \overline{P} \to \ell^{\pm} X$. This is because we use $|A_{\ell^+ X}| = |\overline{A}_{\ell^- X}|$ and $A_{\ell^- X} = \overline{A}_{\ell^+ X} = 0$, which is the case in the SM and in most of its extensions to lowest order in G_F. Thus, in this case, $\lambda_f = 0$.

This source of CP violation can be measured via the asymmetry of "wrong-sign" decays induced by oscillations:

$$\mathscr{A}_{\rm SL}(t) \equiv \frac{\Gamma[\overline{B}(t) \to \ell^+ X] - \Gamma[B(t) \to \ell^- X]}{\Gamma[\overline{B}(t) \to \ell^+ X] + \Gamma[B(t) \to \ell^- X]} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \tag{94}$$

Note that this asymmetry of time-dependent decay rates is actually time independent.

We are now going to give few examples of cases that are sensitive to CP violation that involve mixing.

5.4.1 $B \to \psi K_S$

The "golden mode" with regards to CP violation in interference between mixing and decays is $B \to \psi K_S$. It provides a very clean determination of the angle β of the unitarity triangle.

As we already mentioned we know that, to a very good approximation, in the B system $|q/p| = 1$. In that case we have

$$\lambda_f = e^{-i\phi_B} \frac{\overline{A}_f}{A_f} , \tag{95}$$

where ϕ_B refers to the phase of M_{12} (see Eq. (78)). Within the SM, where the top diagram dominates the mixing, the corresponding phase factor is given by

$$e^{-i\phi_B} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} , \tag{96}$$

to a very good approximation. For $f = \psi K$, which proceeds via a $\bar{b} \to \bar{c}c\bar{s}$ transition, we can write

$$A_{\psi K} = \left(V_{cb}^* V_{cs}\right) T_{\psi K} \tag{97}$$

where $T_{\psi K}$ is the magnitude of the tree amplitude. In principle there is also a penguin amplitude that contributes to the decay. The leading penguin carries the same weak phase as the tree amplitude. The one that carries a different weak phase is highly CKM suppressed and we neglect it. This is a crucial point. Therefore we have

$$|\lambda_{\psi K_S}| = 1, \qquad \mathrm{Im}(\lambda_{\psi K_S}) = \sin 2\beta, \tag{98}$$

to a very good approximation. Thus Eq. (90) applies in this case. We conclude that a measurement of the CP asymmetry in $B \to \psi K_S$ gives a very clean determination of the angle β of the unitarity triangle. Here we were able to overcome QCD by the fact that the decay is dominated by one decay amplitude that cancels in the CP asymmetry. The strong phase arises due to the oscillation and it is related to the known Δm. This CP asymmetry measurement was done and provides at present the most precise determination of any angle or side of the unitarity triangle.

A subtlety arises in this decay that is related to the fact that B^0 decays into ψK^0 while $\overline{B}^0$ decays into $\psi \overline{K}^0$. A common final state, e.g. ψK_S, is reached only via $K^0 - \overline{K}^0$ mixing. We do not elaborate on this point.

There are many more decay modes where a clean measurement of angles can be performed. In your homework, you will work out one more example and even try to get an idea of the theoretical errors.

5.4.2 *K* Decays

CP violation was discovered in K decays, and until recently, it was the only meson where CP violation had been measured. CP violation was first observed in $K_L \to \pi\pi$ decays in 1964, and later in semileptonic K_L decays in 1967. Beside the historical importance, kaon CP violation provides important bounds on the unitarity triangle. Moreover, when we consider generic new physics, CP violation in kaon decays provides the strongest bound on the new physics scale!

While the formalism of CP violation is the same for all mesons, the relevant approximations are different. For the B system, we neglected the width difference and got the very elegant formula, Eq. (90). For the B mesons it is easy to talk in terms of flavour (or CP) eigenstates, and use mass eigenstates only as intermediate states to calculate the time evolutions. For kaons, however, the width difference is very large

$$\frac{\Gamma_S}{\Gamma_L} \sim 600. \tag{99}$$

This implies that, to very good approximation, we can get a pure K_L state: all we have to do is wait. Since we do have a pure K_L state, it is easy to talk in terms of mass eigenstates. Note that it is not easy to get a pure K_S state. At short times we have a mixture of states and, only after the K_S part has decayed, we have a pure K_L state.

In terms of mass eigenstates, CP violation is manifested if the same state can decay to both CP even and CP odd states. This should be clear to you from basic quantum mechanics. Consider a symmetry, that is, an operator that commutes with the Hamiltonian. In the case under consideration, if CP is a good symmetry it implies $[\text{CP}, H] = 0$. When this is the case, any non-degenerate state must be an eigenstate of CP. In a CP-conserving theory, any eigenstate of CP must decay to a state with the same CP parity. In particular, it is impossible to observe a state that can decay to both CP even and CP odd states. Thus, CP violation in kaon decays was established when $K_L \to \pi\pi$ was observed. K_L decays dominantly to three pions, which is a CP odd state. The fact that it decays also to two pions, which is a CP even state, implies CP violation.

We will not get into the details of the calculations, but it should be clear at this stage that the rate of $K_L \to \pi\pi$ must be related to the values of the CKM parameters and, in particular, to its phase. I hope you will find the time to read about it in one of the reviews I mentioned.

Before concluding, we remark on semileptonic CP violation in kaons. When working with mass eigenstates, CP conservation implies that

$$\Gamma(K_L \to \pi^- \ell^+ \nu) = \Gamma(K_L \to \pi^+ \ell^- \bar{\nu}). \tag{100}$$

Experimentally, the above equality was found to be violated, implying CP violation.

In principle, the CP violation in $K_L \to \pi\pi$ and in semileptonic decays are independent observables. Yet, if all the decay amplitudes carry the same phase, these two are related. This is indeed the case in the kaon system, and thus we talk about one parameter that measure kaon CP violation, which is denoted by ε_K. (There is another parameter called ε'_K, but we will not discuss it here.)

5.5 Homework

Question 11 (Condition for CP violation). Using Eq. (77), show that in order to observe CP violation, $\Gamma(B \to f) \neq \Gamma(\overline{B} \to \overline{f})$, we need two amplitudes with different weak and strong phases.

Question 12 (Mixing formalism). In this question, you are asked to develop the general formalism of meson mixing.

1. Show that the mass and width differences are given by

$$4(\Delta m)^2-(\Delta\Gamma)^2=4(4|M_{12}|^2-|\Gamma_{12}|^2), \qquad \Delta m\Delta\Gamma=4\mathrm{Re}(M_{12}\Gamma_{12}^*), \tag{101}$$

   and that

$$\left|\frac{q}{p}\right| = \left|\frac{\Delta m - i\Delta\Gamma/2}{2M_{12} - i\Gamma_{12}}\right|. \tag{102}$$

2. When CP is a good symmetry all mass eigenstates must also be CP eigenstates. Show that CP invariance requires

$$\left|\frac{q}{p}\right| = 1. \tag{103}$$

3. In the limit $\Gamma_{12} \ll M_{12}$ show that

$$\Delta m = 2|M_{12}|, \qquad \Delta\Gamma = 2|\Gamma_{12}|\cos\theta, \qquad \left|\frac{q}{p}\right| = 1. \tag{104}$$

4. Derive Eqs. (88).
5. Derive Eq. (90).
6. Show that when $\Delta\Gamma = 0$ and $|q/p| = 1$

$$\begin{aligned}\Gamma(B \to X\ell^-\bar{\nu})[t] &= e^{-\Gamma t}\sin^2(\Delta m t/2),\\ \Gamma(B \to X\ell^+\nu)[t] &= e^{-\Gamma t}\cos^2(\Delta m t/2).\end{aligned} \tag{105}$$

Question 13 ($B \to \pi^+\pi^-$ and CP violation). One of the interesting decays to consider is $B \to \pi\pi$. Here we briefly discuss it.

1. First assume that there is only a tree-level decay amplitude (that is, neglect penguin amplitudes). Draw the Feynman diagram of the amplitude, paying special attention to its CKM dependence.
2. In that case, which angle of the unitarity triangle is the time dependent CP asymmetry, Eq. (90), sensitive to?
3. Can you estimate the error introduced by neglecting the penguin amplitude? (Note that one can use isospin to reduce this error. Again, you are encouraged to read about it in one of the reviews.)

Question 14 (B decays and CP violation). Consider the decays $\bar{B}^0 \to \psi K_S$ and $\bar{B}^0 \to \phi K_S$. Unless explicitly noted, we always work within the framework of the standard model.

1. $\bar{B}^0 \to \psi K_S$ is a tree-level process. Write down the underlying quark decay. Draw the tree-level diagram. What is the CKM dependence of this diagram? In the Wolfenstein parametrization, what is the weak phase of this diagram?
2. Write down the underlying quark decay for $B^0 \to \phi K_S$. Explain why there is no tree-level diagram for $B^0 \to \phi K_S$.
3. The leading one-loop diagram for $B^0 \to \phi K_S$ is a gluonic penguin diagram. As we have discussed, there are several diagrams and only their sum is finite. Draw a representative diagram with an internal top quark. What is the CKM dependence of the diagram? In the Wolfenstein parametrization, what is the weak phase of the diagram?
4. Next we consider the time dependent CP asymmetries. We define as usual

$$\lambda_f \equiv \frac{\bar{A}_f}{A_f}\frac{q}{p}, \qquad A_f \equiv A(B^0 \to f), \qquad \bar{A}_f \equiv A(\bar{B}^0 \to f). \tag{106}$$

   In our case we neglect subleading diagrams and then we have $|\lambda_f| = 1$ and thus

$$a_f \equiv \frac{\Gamma(\bar{B}^0(t) \to f) - \Gamma(B^0(t) \to f)}{\Gamma(\bar{B}^0(t) \to f) + \Gamma(B^0(t) \to f)} = -\mathrm{Im}\lambda_f \, \sin(\Delta m_B \, t) \tag{107}$$

   Both $a_{\psi K_S}$ and $a_{\phi K_S}$ measure the same angle of the unitarity triangle. That is, in both cases, $\mathrm{Im}\lambda_f = \sin 2x$ where x is one of the angles of the unitarity triangle. What is x? Explain.
5. Experimentally,

$$\mathrm{Im}\lambda_{\psi K_S} = 0.68(3), \qquad \mathrm{Im}\lambda_{\phi K_S} = 0.47(19). \tag{108}$$

   Comment about these two results. In particular, do you think these two results are in disagreement?
6. Assume that in the future we will find

$$\mathrm{Im}\lambda_{\psi K_S} = 0.68(1), \qquad \mathrm{Im}\lambda_{\phi K_S} = 0.32(3). \tag{109}$$

That is, that the two results are not the same. Below are three possible "solutions". For each solution explain if you think it could work or not. If you think it can work, show how. If you think it cannot, explain why.

a. There are standard model corrections that we neglected.
b. There is a new contribution to $B^0 - \bar{B}^0$ mixing with a weak phase that is different from the SM one.
c. There is a new contribution to the gluonic penguin with a weak phase that is different from the SM one.

Question 15 (Decay of mass eigenstates). Derive Eq. (91). The idea is to understand that when we talk about mass eigenstates, we are talking about "late times," $t \gg x\Gamma$ so that the $\sin(\Delta m t)$ term can be averaged out.

6 Putting It All Together: The Big Picture

After we got a taste of how to probe the flavour sector, we are ready to ask: what are the results? That is, how compatible are all the measurements? As we explained, in principle we need four measurements to define the flavour parameters and the rest are checks on the model. Then we can ask what are the implications of these results on the big question of physics, that is, what is the fundamental Lagrangian of Nature.

6.1 The Current Status of the SM Flavour Sector

Out of the four flavour parameters of the SM, two are known to high accuracy, and there is a very good agreement between the various ways they are determined. These two parameters are λ and A of the Wolfenstein parametrization of the CKM matrix. Thus, it is customary to plot all the other measurements as bounds on the rescaled unitarity triangle, that depend only on two parameters, ρ and η. There are many measurements and bounds and within the errors each of them gives an allowed range in the $\rho - \eta$ plane for the one undetermined apex of the rescaled unitarity triangle.

What I am about to show you (in the next page, please do not peek!) is the most recent compilation of all these results. I am sure you have seen it before. Yet, I hope that now you appreciate it much more. In class, I always stop at this point. I like to make sure the students do not miss the big moment, and that they see how amazing physics is. If I knew how to play a trumpet, this would be the moment that I would use it. In a written version, it is a bit harder. I cannot stop you from looking at the next page and I certainly cannot play the trumpet here. Still I do ask you to take a break here. Make sure you fully understand what you are going to see.

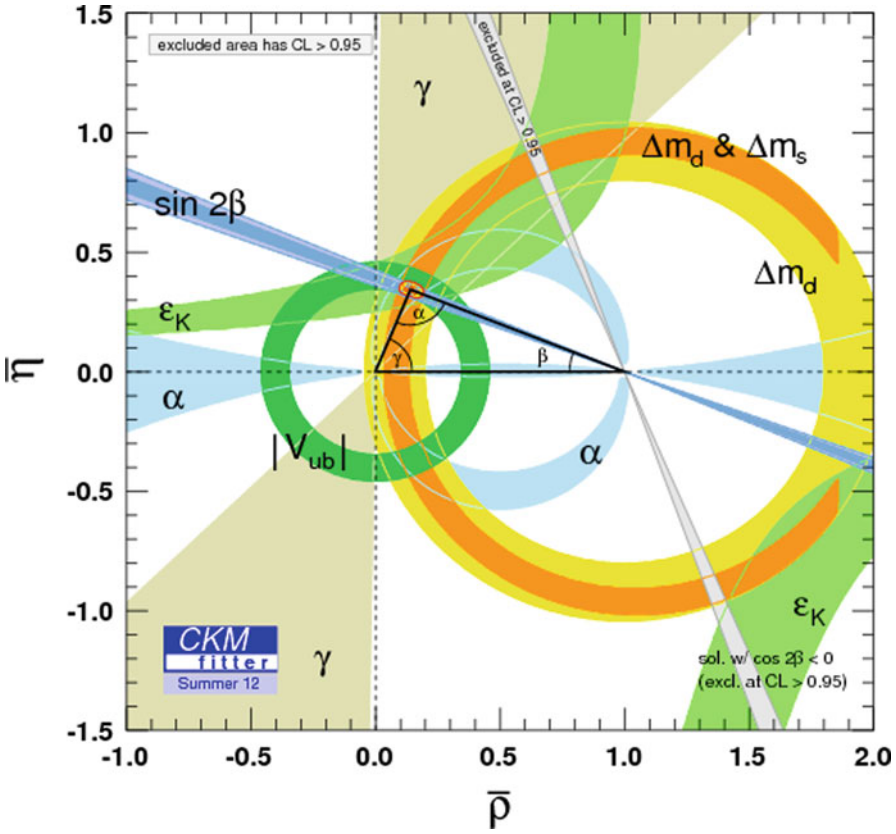

Fig. 3 Global fit to the unitarity triangle based on all available data [17]

Now, that you are ready, take your time and look at Fig. 3. What you see are many bounds all overlapping at one small area in the $\rho - \eta$ plane. That is, taking the two measurements as determining ρ and η, all the rest are checks on the SM. You see that the flavour sector of the SM passes all its tests. Basically, all the measurements agree.

The most important implication of this triumph of theoretical and experimental physics is the following statement: The Cabibbo-Kobayashi-Maskawa mechanism is the dominant source of flavour and CP violation in low-energy flavour-changing processes. This is a very important statement. Indeed the Nobel prize was awarded to Kobayashi and Maskawa in 2008 because it is now experimentally proven that the KM phase is the one which explains the observed CP violation in Nature.

6.2 Instead of a Summary: The New Physics Flavour Problem

The success of the SM can be seen as a proof that it is an effective low energy description of Nature. There are, however, many reasons to suspect that the SM has to be extended. A partial list includes the hierarchy problem, the strong CP problem, baryogenesis, gauge coupling unification, the flavour puzzle, neutrino masses, and gravity. We are therefore interested in probing the more fundamental theory. One way to go is to search for new particles that can be produced at yet unreached energies. Another way to look for new physics is to search for indirect effects of heavy unknown particles. Flavour physics is used to probe such indirect signals of physics beyond the SM.

In general, flavour bounds provide strong constraints on new physics models. This fact is called "the new physics flavour problem". The problem is actually the mismatch between the new physics scale that is required in order to solve the hierarchy problem and the one that is needed in order to satisfy the experimental bounds from flavour physics.

In order to understand what the new physics flavour problem is, let us first recall the hierarchy problem. In order to prevent the Higgs mass from getting a large radiative correction, new physics must appear at a scale that is a loop factor above the weak scale

$$\Lambda \lesssim 4\pi m_W \sim 1 \text{ TeV}. \tag{110}$$

Here, and in what follows, Λ represents the new physics scale. Note that such TeV new physics can be directly observed in collider searches.

While the SM scalar sector is "unnatural", its flavour sector is impressively successful.[14] This success is linked to the fact that the SM flavour structure is special. As we already mentioned, the charged current interactions are universal (in the mass basis this is manifest through the unitarity of the CKM matrix) and FCNCs are highly suppressed: they are absent at the tree level and at the one-loop level they are further suppressed by the GIM mechanism. These special features are important in order to explain the observed pattern of weak decays. Thus, any extension of the SM must conserve these successful features.

Consider a generic new physics model, where the only suppression of FCNCs processes is due to the large masses of the particles that mediate them. Naturally, these masses are of the order of the new physics scale, Λ. Flavour physics, in particular measurements of meson mixing and CP violation, puts severe constraints on Λ. In order to find these bounds we take an effective field theory approach. At the weak scale we write all the non-renormalizable operators that are consistent with the

[14] The flavour structure of the SM is interesting since the quark masses and mixing angles exhibit some hierarchies. These are not explained within the SM, and this fact is usually called "the SM flavour puzzle". This puzzle is different from the new physics flavour problem that we are discussing here.

gauge symmetry of the SM. In particular, flavour-changing four Fermi operators of the form (the Dirac structure is suppressed)

$$\frac{q_1\bar{q}_2 q_3\bar{q}_4}{\Lambda^2}, \tag{111}$$

are allowed. Here q_i can be any quark as long as the electric charges of the four fields in Eq. (111) sum up to zero. We emphasize that there is no exact symmetry that can forbid such operators. This is in contrast to operators that violate baryon or lepton number that can be eliminated by imposing symmetries like $U(1)_{B-L}$ or R-parity. The strongest bounds are obtained from meson mixing and CP-violation measurements. Depending on the mode we find bounds of the order

$$\Lambda \gtrsim \text{few} \times 10^4 \text{ TeV}. \tag{112}$$

There is tension between the new physics scale that is required in order to solve the hierarchy problem, Eq. (110), and the one that is needed in order not to contradict the flavour bounds, Eq. (112). The hierarchy problem can be solved with new physics at a scale $\Lambda \sim 1\,\text{TeV}$. Flavour bounds, on the other hand, require $\Lambda \gtrsim 10^4\,\text{TeV}$. This tension implies that any TeV scale new physics cannot have a generic flavour structure. This is the new physics flavour problem.

Flavour physics has been mainly an input to model building, not an output. The flavour predictions of any new physics model are not a consequence of its generic structure but rather of the special structure that is imposed to satisfy the severe existing flavour bounds. It is clearly a very interesting open question to determine the New Physics model and how it deals with flavour.

6.3 Concluding Remarks

This is a good time to finish the lectures. I hope that you gained some understanding of flavour physics, how it was used to shape the SM as we know it, and why it is so important in our quest to find the theory that extends the SM. In the near future we expect more data in the energy frontier, as well as more flavour data. It can be really fun to see how the two can work together to show us what Nature is at really short distances, that is, to help us in getting a better answer to the fundamental question of physics

$$\mathscr{L} = ? \tag{113}$$

Acknowledgements I thank Joshua Berger, Mario Martone, and Dean Robinson for comments on the manuscript. The work of YG is supported by NSF grant number PHY-0757868 and by a U.S.-Israeli BSF grant.

References

1. G. Isidori, Flavor physics and CP violation, in *2012 European School of High-Energy Physics*, Anjou, 6–19 Jun 2012. To be published in the School proceedings, arXiv:1302.0661 [hep-ph].
2. O. Gedalia and G. Perez, TASI 2009 lectures – flavor physics, in *TASI 2009*, Boulder. CO, USA, arXiv:1005.3106 [hep-ph].
3. G. Isidori, Y. Nir and G. Perez, Flavor physics constraints for physics beyond the standard model. Ann. Rev. Nucl. Part. Sci. **60**, 355 (2010) [arXiv:1002.0900 [hep-ph]].
4. G. Isidori, B physics in the LHC era, in *65th Scottish Universities Summer School in Physics*, St. Andrews. arXiv:1001.3431 [hep-ph].
5. A. J. Buras, Flavour Theory: 2009. PoS EPS **HEP2009**, 024 (2009) [arXiv:0910.1032 [hep-ph]].
6. U. Nierste, Three lectures on meson mixing and CKM phenomenology, in *Helmholtz International Summer School on Heavy quark physics*, Dubna. arXiv:0904.1869 [hep-ph].
7. M. Artuso, E. Barberio and S. Stone, *B* meson decays. PMC Phys. A **3**, 3 (2009) [arXiv:0902.3743 [hep-ph]].
8. Y. Nir, Probing new physics with flavor physics (and probing flavor physics with new physics), in *PiTP 2007 on Standard Model and Beyond*, IAS, Princeton. arXiv:0708.1872 [hep-ph].
9. A. Hocker and Z. Ligeti, CP violation and the CKM matrix. Ann. Rev. Nucl. Part. Sci. **56**, 501 (2006) [arXiv:hep-ph/0605217].
10. M. Neubert, Effective field theory and heavy quark physics, in *TASI 2004*, Boulder. CO, USA, arXiv:hep-ph/0512222.
11. Y. Nir, CP violation in meson decays, in *Third CERN-CLAF School of High Energy Physics*, Malargue. arXiv:hep-ph/0510413.
12. A. J. Buras, Flavour physics and CP violation, in *CERN School* 2004. arXiv:hep-ph/0505175.
13. Z. Ligeti, Introduction to heavy meson decays and CP asymmetries, in *Proceedings of 30th SLAC Summer Institute on Particle Physics: Secrets of the B Meson (SSI 2002)*, SLAC, Menlo Park, 5–16 Aug 2002, pp. L02 [arXiv:hep-ph/0302031].
14. I. I. Y. Bigi and A. I. Sanda, CP violation. Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **9**, 1 (2000).
15. G. C. Branco, L. Lavoura and J. P. Silva, CP violation. Int. Ser. Monogr. Phys. **103**, 1 (1999).
16. J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012). Recent updates can be found online at pdg.lbl.gov.
17. J. Charles et al., CKMfitter Group Collaboration, Eur. Phys. J. C **41**, 1 (2005) [hep-ph/0406184]; updated results and plots available at: http://ckmfitter.in2p3.fr

Beyond the Standard Model Phenomenology and the ElectroWeak Symmetry Breaking

Giacomo Cacciapaglia

Abstract This report contains a review of the motivations and main directions in the construction of models beyond the standard model. The lectures follow the role played by symmetries in model building as a guiding line. After a review of the symmetries of the Brout-Englert-Higgs sector of the standard model, and the role of spin in the naturalness argument, a pedagogical introduction to supersymmetry and extra dimensions is given, both being seen as an extension of Poincarè symmetries of space-time.

1 Why Do We Need to Go Beyond the Standard Model (BSM)?

The Standard Model of Particle Physics (SM for brevity) was proposed in the 1960s and early 1970s by Sheldon Glashow [26], Stephen Weinberg [46] and Abdus Salam to describe, in terms of fundamental degrees of freedom (or particles), the theory of electro-weak interactions first proposed by Enrico Fermi in the 1930s. The effective 4-fermion interactions in Fermi's theory are replaced by the presence of massive gauge bosons, the charged $W^{\pm}$ and the neutral Z: while the W is required to explain the beta decay of atoms, the Z is a prediction of the gauge structure introduced in the model and its discovery in 1984 was a crucial validation of the theory. The model can be divided in the following sectors:

- $SU(3)_c$ gauge group (colour), that describes strong interactions (QCD);
- $SU(2)_L \times U(1)_Y$ that describes electro-weak interactions, it is spontaneously broken to an exact $U(1)_{em}$ gauge group describing electromagnetic interactions (QED);

G. Cacciapaglia (✉)
Université de Lyon, F-69622 Lyon, France
Université Lyon 1, Villeurbanne, France
CNRS/IN2P3, UMR5822, Institut de Physique Nucléaire de Lyon, F-69622 Villeurbanne Cedex, France
e-mail: g.cacciapaglia@ipnl.in2p3.fr

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_3

– The Brout-Englert-Higgs (BEH) sector, that describes the breaking of the electro-weak symmetry (EWSB) by means of a scalar field that develops a vacuum expectation value;
– Fermionic matter: the model contains three generations of fermions, each one consisting of two quarks (that describe the baryonic matter, like pions and nucleons), a charged lepton (electron) and a neutrino;
– Yukawa interactions between fermions and the Higgs sector that give mass to the fermions and are responsible for flavour mixing.

The Standard Model is therefore composed by sectors, or building blocks, which are closely related to each other and their non trivial interplay is essential for the success of the model to describe phenomena observed at various energies. What glues together and inter-links these building blocks? The apparently complex structure of the Standard Model is based on symmetries! Gauge symmetries describe the behaviour and interactions of vector bosons, the force carriers. The generation of masses is based on a mechanism of spontaneous symmetry breaking: the theory is invariant under the $SU(2)_L \times U(1)_Y$ gauge symmetry, which forbids masses for both the gauge bosons and fermions in the SM. However the vacuum of the scalar BEH field is not invariant! The field content of each generation of fermions is fixed by the cancellation of gauge anomalies in the theory: all gauge symmetries of the SM, in fact, play a role in determining the number and charges of fields in a single generation. A different choice from the one observed in nature would invalidate the theory by breaking explicitly one or more gauge symmetries.

The presence of three generations is not fully understood yet. Nevertheless, three is the minimal number of generations that allows for non-trivial phases in the Yukawa sector. Such phases break the invariance under the CP symmetry (Charge conjugation and Parity), and this realisation was worth the Nobel prize for Kobayashi and Maskawa [33]. Furthermore, the structure of Yukawa couplings, which is determined by the quantum numbers of the fermions, ensures the presence of two unbroken global symmetries – Baryon and Lepton number conservation. They are accidental symmetries, in the sense that they have not been used in the construction of the model. It is well known that both Baryon and Lepton symmetries are broken by non-perturbative effects, however the combination $B - L$ is exactly conserved, and this is enough to ensure the stability of the proton.

The interactions of the SM have been tested very accurately in the last 30 years, mainly by LEP (Large Electron Positron collider) at CERN and the proton-antiproton collider Tevatron at Fermilab. The data show in fact an amazing agreement with the SM predictions (including quantum loop effects), with some measurements agreeing up to a part in 1,000 [5]. As of today, the sector of the theory which is the least experimentally known is the EWSB sector, i.e. the physics associated with the BEH field. In the Standard Model, this sector is described in terms of a complex scalar doublet of $SU(2)_L$. After the EWSB, 3 of the degrees of freedom of this field are eaten by the massive gauge bosons $W^{\pm}$ and Z providing them with a longitudinal polarisation, while the fourth is a physical scalar, the Higgs

boson.[1] The Higgs boson has eluded all experimental efforts to discover it,[2] until July 2012 when the LHC collaborations CMS and ATLAS announced the discovery of a new resonance at a mass of 125 GeV [1, 14], that has similar properties as the Higgs boson (see Bill Murray's contribution in these proceedings). This discovery seems to complete the validation of the Standard Model, as all of the predicted particles and interactions have now been observed. So, why do we still need to talk about New Physics?

1.1 Evidences and Hints of New Physics (BSM)

There are several evidences and hints that seem to suggest the presence of New Physics: they can be roughly divided into two classes. On one hand, there are experimental results (**EXP**) that cannot be explained within the Standard Model. On the other, purely theoretical considerations (**TH**) do not allow us to accept the SM as the ultimate theory. Here it follows a brief incomplete list:

- **EXP**: *neutrino masses.* The observation and patterns of neutrino oscillations [38] suggest that at least two of the three neutrino species have mass, and from cosmological observations we know that the mass scale has to be very small (sub-eV). The masses can be added by simply extending the SM with three right-handed singlet neutrinos. See-saw mechanisms [45] would hint that the mass of the new states is at the $10^{3\div 13}$ GeV level.
- **EXP**: *Dark Matter in the Universe.* Twenty-three percent of the total mass of our universe [2, 31] is made of non-baryonic and non-luminous matter, which is not accounted for in the SM. This observation is also supported by astrophysical observations: rotation curves of disk galaxies, gravitational micro-lensing of galaxy clusters and large structure formation models. A particle interpretation would suggest a weakly interacting particle with a mass of $\sim 100 \div 1{,}000$ GeV. This new particle must be neutral and stable (on cosmological time scales).
- **EXP**: *Baryon asymmetry in the Universe.* The Universe is populated by baryons, however the number of anti-baryons is very scarce. To explain this, the mechanism of baryogenesis has been proposed: it requires three conditions formulated by Sakharov [41], namely that baryon symmetry is broken (by anomalies in

[1] There has been a dispute on the proper name this boson should have. In the author's view, Peter Higgs was the first to explicitly mention the existence and properties of such particle [29, 30], therefore it deserves its name. On the other hand, the mechanism of EWSB by a vacuum expectation value has been applied to gauge symmetries by Brout and Englert in an earlier paper [24], and also in a publication by Guralnik, Hagen and Kibble later the same year [28]. For this reasons, the scalar field and mechanism are dubbed BEH, while the scalar particle is named Higgs boson in these proceedings.

[2] In a 1976 paper [23], John Ellis and collaborators pointed out how difficult it is to discover such a state, to the point of discouraging any experimental effort in such directions. Fortunately, their advice has not been followed.

the SM), that the model violates CP (Yukawa interactions in the SM) and that there is a non-thermal process (the EW phase transition in the SM). However, in the SM, the amount of CP violation in the quark sector is not enough to explain the baryon density at present days by many orders of magnitude.

- **TH/EXP**: *gauge coupling unification.* Running the SM gauge couplings at high energies, their values tend to converge to the same value at $\sim 10^{15\div 17}$ GeV. A new unified gauge theory (GUT) may be present at such energy scales [39].
- **TH**: *quantum gravity.* Classical gravity, well described by general relativity, should break down at energy scales close to the Planck mass $M_{Pl} \sim 10^{19}$ GeV, where quantum effects should arise. The SM is necessarily invalidated at such energy. This scale can be lowered in models where the fundamental Planck scale is lower (for instance, models with extra dimensions [9]).
- **TH**: *Higgs mass (electro-weak scale) instability.* The Higgs mass is sensitive to quadratically divergent radiative corrections, thus it is unstable. In physical terms, new states that couple to the Higgs will generate corrections to its mass which are proportional to the mass of the new state. As the mass of the Higgs cannot exceed $\sim$1 TeV (we now know that it may be 125 GeV), such states must be at the TeV scale in order to avoid large cancellations with the tree level mass (naturalness argument).

All the entries in this non-exhaustive list point to the presence of new phenomena or new particles at a given scale. It is striking that only two entries require new particles at of below the TeV scale (which is accessible at the LHC): a weakly interacting particle candidate for Dark Matter and the naturalness argument on the Higgs mass. Other cases can also be lowered to the TeV scale: TeV scale see saw in the case of neutrino masses, accelerated unification or gravity in extra dimensions (TeV scale Black Holes); however, this is not required! In these lectures we will be mainly interested in the naturalness argument which involves the BEH sector of the Standard Model (and, sometimes, Dark Matter). Before starting our journey in the landscape of physics Beyond the Standard Model (BSM), it is important to point out what the naturalness argument really is: it is a theoretical prejudice against the Standard Model as the fundamental theory of particle physics. In fact, the Standard Model is based upon a renormalisable lagrangian, therefore divergences can be consistently reabsorbed in tree level parameters (the BEH field mass in this case) and, no matter how large they are, they leave no trace in observable quantities! The naturalness argument bases its power on the assumption that the SM is an effective model valid up to a cutoff, i.e. a high energy scale above which the model must be replaced by a more fundamental theory or where new particles should be included to complete the model. The truth told, we do know that above the Planck scale a theory of quantum gravity must replace the SM! One may conclude that a natural Higgs boson must have a mass close to the Planck scale (and thus goes the W and the Z bosons), unless a protection mechanism is at work. Protection mechanisms in particle physics are called **symmetries**. A central point in these lectures will be the role of symmetries in models of physics Beyond the Standard Model.

2 The EWSB Sector, and the Role of Symmetries

In the SM, the electro-weak symmetry breaking (EWSB) is *described* by the Brout-Englert-Higgs scalar field. Such field is a complex scalar (spin-0), which transforms as a doublet under the $SU(2)_L$ symmetry and has hypercharge $1/2$ (and no colour). The most general renormalisable action for the BEH scalar ϕ is:

$$\mathscr{S}_{BEH} = \int d^4x \; (D^\mu \phi)^\dagger D_\mu \phi - m_\phi^2 \phi^\dagger \phi - \frac{\lambda}{2}(\phi^\dagger \phi)^2 \,; \tag{1}$$

where

$$\phi = \begin{pmatrix} \varphi^+ \\ \varphi_0 \end{pmatrix} \tag{2}$$

is the doublet in components, and $\phi^\dagger \phi = \varphi^- \varphi^+ + \varphi_0^* \varphi_0$. The above action is invariant under local $SU(2)_L \times U(1)_Y$ transformations, provided that the covariant derivative D_μ contains the proper gauge fields

$$D_\mu \phi = \left(\partial_\mu - i g \sum_{a=1}^{3} W_\mu^a t^a - i g' \frac{1}{2} B_\mu \right) \phi \,. \tag{3}$$

You should be already very familiar with the fact that, for $m_\phi^2 < 0$, the potential for the BEH field in Eq. (1) has minima with $\langle \phi \rangle \neq 0$. In fact, the equation of motion for a static ϕ (i.e. such that $\partial_\mu \phi = 0$) is:

$$\left(-m_\phi^2 - \lambda \phi^\dagger \phi \right) \phi = 0 \,. \tag{4}$$

The solutions are $\langle \phi \rangle = 0$, which is a maximum, and

$$\langle \phi^\dagger \phi \rangle = -\frac{m_\phi^2}{\lambda} \tag{5}$$

which is a minimum of the potential. The theory is still invariant under $SU(2)_L \times U(1)_Y$ gauge symmetries: in fact both the action in Eq. (1) and the minimum condition in Eq. (5) are invariant, thus we can use $SU(2)_L \times U(1)_Y$ transformations to write the solution as

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} , \quad v^2 = -2 \frac{m_\phi^2}{\lambda} \,. \tag{6}$$

The explicit solution in Eq. (6) is NOT invariant under the gauge symmetries, thus, when plugged back into Eq. (1), it will give mass to 3 of the gauge bosons:

$$m_{W\pm}^2 = \frac{g^2}{4}v^2\,, \qquad m_Z^2 = \frac{g^2+g'^2}{4}v^2\,. \tag{7}$$

Therefore, we can say that it is not the condition $m_\phi^2 < 0$ that breaks the symmetry, but the choice of an explicit form of the non-trivial vacuum solution for the BEH field. Now, 3 out of the 4 degrees of freedom in the BEH field are eaten up by the massive $W^\pm$ and Z to play the role of the longitudinal polarisation of the massive vector field (which is absent in the massless limit), and the remaining one appears in the spectrum as a massive physical scalar:

$$\phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}\,, \tag{8}$$

where the field h in this expansion is the Higgs boson, and it has a mass

$$m_h^2 = -2m_\phi^2 = \lambda v^2\,. \tag{9}$$

Why is this a mere *description* of the EWSB? The mechanism relies on the fact that the mass has the wrong sign, i.e. $m_\phi^2 < 0$, however there is no explanation of what the origin of such negative mass is! Explaining the origin of the negative mass square can be addressed in BSM models.

2.1 Symmetries: Exposed and Hidden Ones

Let's have a closer look at the BEH action in Eq. (1), reported below:

$$\mathscr{S}_{BEH} = \int d^4x\ (D^\mu\phi)^\dagger D_\mu\phi - m_\phi^2\phi^\dagger\phi - \frac{\lambda}{2}(\phi^\dagger\phi)^2\,. \tag{10}$$

This action was built based on invariance under the SM gauge symmetries: it is the most general gauge invariant and renormalisable lagrangian for a scalar field that transforms as a doublet of SU(2). Is it there any other symmetry that we missed?

There are three kinds of symmetries:

1. Symmetries of space time: Poincaré (rotations, translations, Lorentz boosts) and C, P and T.[3]
2. Gauge (local) symmetries: $SU(2)_L \times U(1)_Y$ (the colour symmetry $SU(3)_c$ plays no role here).

[3] C P and T are discrete symmetries of space-time: C stands for charge inversion, P for spatial parity and T for time inversion. All field theories are, by construction, CPT invariant, i.e. invariant under a combination of the three discrete symmetries.

3. Accidental (global) symmetries.
4. Discrete symmetries.

The action in Eq. (10) contains a very important hidden (and approximate) global symmetry: the **custodial symmetry**. Let's for a moment ignore gauge interactions: if we do so, the action only depends on the combination $\phi^\dagger\phi$ (and $(\partial^\mu\phi^\dagger)\partial_\mu\phi$). This element can be written as

$$\phi^\dagger\phi = \varphi^-\varphi^+ + \varphi_0^*\varphi_0 = \frac{1}{2}\left(\varphi_1^2 + \varphi_2^2 + \varphi_3^3 + \varphi_4^2\right) \tag{11}$$

where $\varphi^+ = \frac{\varphi_1 + i\varphi_2}{\sqrt{2}}$ and $\varphi_0 = \frac{\varphi_3 + i\varphi_4}{\sqrt{2}}$ are decomposed in terms of real and imaginary parts. If we define a real 4-component field

$$\tilde{\Phi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{pmatrix}, \tag{12}$$

the action in Eq. (10) can be re-written as

$$\mathscr{S}_{BEH} = \int d^4x \, \frac{1}{2}(\partial^\mu\tilde{\Phi}^T)\partial_\mu\tilde{\Phi} - \frac{1}{2}m_\phi^2\tilde{\Phi}^T\tilde{\Phi} - \frac{\lambda}{8}(\tilde{\Phi}^T\tilde{\Phi})^2 \,. \tag{13}$$

Now, it is evident that the action is invariant under an orthogonal rotation of the 4-component real vector $\tilde{\Phi}$, i.e. the action is invariant under SO(4) $\sim$ SU(2)×SU(2)! This hidden global symmetry is at the origin of the so-called custodial symmetry!

The 4-dimensional rotation group SO(4) is equivalent to SU(2)×SU(2): the latter structure becomes evident if we rewrite the BEH field as a 2×2 matrix in the following way:

$$\Phi = \begin{pmatrix} \varphi_0^* & \varphi^+ \\ -\varphi^- & \varphi_0 \end{pmatrix}. \tag{14}$$

This field transforms as a bi-doublet under $SU(2)_L \times SU(2)_R$, i.e.

$$\Phi \to U_L \Phi U_R^\dagger$$

where $U_{L/R}$ are two independent SU(2) transformations. The action in Eq. (10) can now be re-written as

$$\mathscr{S}_{EBH} = \int d^4x \, \frac{1}{2}\mathrm{Tr}(\partial^\mu\Phi^\dagger)\partial_\mu\Phi - \frac{1}{2}m_\phi^2\mathrm{Tr}\Phi^\dagger\Phi - \frac{\lambda}{8}(\mathrm{Tr}\Phi^\dagger\Phi)^2 \,, \tag{15}$$

where $\mathrm{Tr}\Phi^\dagger\Phi$ is the matricial trace. This latter lagrangian is explicitly invariant under a $SU(2)_L \times SU(2)_R$ transformation:

$$\mathrm{Tr}\Phi^\dagger\Phi \to \mathrm{Tr}U_R\Phi^\dagger U_L^\dagger U_L \Phi U_R^\dagger = \mathrm{Tr}\Phi^\dagger\Phi U_R^\dagger U_R = \mathrm{Tr}\Phi^\dagger\Phi\,. \tag{16}$$

In this notation, the vacuum solution in Eq. (6) can be written as

$$\langle\Phi\rangle = \frac{v}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{17}$$

This vacuum is invariant under a SU(2) symmetry, defined as the symmetry for which $U_L = U_R$:

$$\langle\Phi\rangle \to \langle U_L\Phi U_R^\dagger\rangle = \frac{v}{\sqrt{2}}U_L U_R^\dagger = \langle\Phi\rangle\,. \tag{18}$$

In other words, the Higgs vacuum is breaking $SU(2)_L \times SU(2)_R \to SU(2)_D$! The preserved $SU(2)_D$ is the custodial symmetry. This global symmetry is very important in the SM: in fact, it protects the relative values of the W and Z mass against radiative corrections. To understand this fact, we need to include gauge interactions, that were put aside at the beginning of the section: the SM gauge group is not $SU(2)_L \times SU(2)_R$, however the $SU(2)_L$ is same while the $U(1)_Y$ can be identified with the diagonal generator of $SU(2)_R$. In other words, the $SU(2)_R$ is only partially gauged, and this fact explicitly violates it. If we imagine a world where both SU(2) were gauged, then the BEH vacuum solution would contain 3 massless gauge bosons (corresponding to the $SU(2)_D$) and 3 massive ones with equal mass:

$$m'_W = m'_Z = \frac{\sqrt{g^2 + g'^2}\,v}{2}\,. \tag{19}$$

In the SM, the $SU(2)_R$ is missing its charged gauge bosons: therefore, the massless W disappears, and the mass of the massive one would be given by the above formula, without the g'. This structure of the W and Z mass is therefore dictated by the global symmetries of the BEH field, and this approximate symmetry will tend to protect the ratio of the two masses from large corrections.

This fact allows to define a parameter

$$\rho = \frac{m_W^2}{m_Z^2\cos^2\theta_W}\,, \qquad \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}\,. \tag{20}$$

At three level, ρ is exactly 1. As the custodial symmetry is not an exact symmetry in the SM, radiative corrections tend to spoil the equality, however, due to the custodial symmetry, the corrections to ρ are of the order of 10^{-3}! Keeping such corrections small in models of new physics is a big challenge!

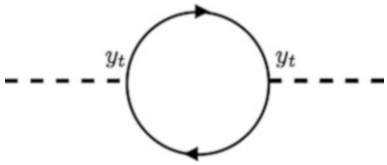

Fig. 1 Top loop contribution to the Higgs mass

2.2 Radiative Stability

The Higgs mass (and thus the electro-weak scale) suffers from divergent radiative corrections. To better understand this statement, let's closely look at the contribution from the top quark. In the SM, fermion masses are generated via Yukawa interactions. For the top

$$\mathscr{L}_{top} = -y_t \bar{Q} \phi t_R + h.c. \tag{21}$$

where $\bar{Q} = (\bar{t}_L, \bar{b}_L)$ is the left-handed $SU(2)_L$ doublet and t_R the right-handed singlet. This interaction will generate a top mass $m_{top} = \frac{y_t v}{\sqrt{2}}$, once the BEH field is replaced by its vacuum solution. Similar interactions can be added to generate a mass for the bottom, and for the other quarks and leptons. In the following, we will focus uniquely on the top because it is the only quark with large coupling to the Higgs: in fact, numerically $y_t \sim 1$.

The above Yukawa interaction will also contribute to the ϕ mass via the loop diagram in Fig. 1:

$$\begin{aligned}
-i\delta m_\phi^2 &= -3(iy_t)^2 \int \frac{d^4k}{(2\pi)^4} \,\mathrm{Tr} \frac{ik^\mu \gamma_\mu}{k^2} \frac{1+\gamma^5}{2} \frac{ik^\nu \gamma_\nu}{k^2} \frac{1-\gamma^5}{2} \\
&= -\frac{3y_t^2}{2(2\pi)^4} \int d^4k \, \frac{\mathrm{Tr} k^\mu \gamma_\mu \, k^\nu \gamma_\nu \, (1-\gamma^5)}{k^4} \\
&= -\frac{3y_t^2}{2(2\pi)^4} \int d^4k \, 4\frac{1}{k^2} \\
&= \frac{3y_t^2}{8\pi^4} \, i 2\pi^2 \int k_E dk_E \\
&= i \frac{3y_t^2}{8\pi^2} \int dk_E^2
\end{aligned} \tag{22}$$

The last integral is clearly divergent. One way to regularise the divergence is to add a hard cutoff to the integral, i.e. $\int dk_E^2 \to \int_0^{\Lambda^2} dk_E^2 = \Lambda^2$. Thus, the corrected ϕ mass would be

$$m_\phi^2\Big|_{1-\text{loop}} = m_\phi^2 - \frac{3y_t^2}{8\pi^2}\Lambda^2\,. \tag{23}$$

Equation (23) contains two intriguing features: first of all, the contribution of the top loop is negative, thus this may be the source of the EWSB which was apparently missing in the standard model. In other words, if we start with vanishing BEH mass, $m_\phi^2 = 0$, the top quantum corrections will generate a negative contribution. The second feature is related to the size of such a contribution: in fact, the Higgs boson mass is related to the BEH mass, and we now know that the Higgs is light with a mass of 125 GeV, therefore one can estimate the value the cutoff Λ should have:

$$m_h^2 = -2m_\phi^2 = \frac{3y_t^2}{4\pi^2}\Lambda^2 = (125\text{ GeV})^2 \Rightarrow \Lambda = 450\,\text{GeV}\,. \tag{24}$$

This implies that the divergent loop should be cut off by New Physics at a scale not far from the TeV, which is many orders of magnitude below the natural scale where we would have expected the SM to lose its validity: the Planck mass 10^{19} GeV. This offset of 16 orders of magnitude is called the **hierarchy problem**. There is a "weaker" hierarchy problem, dubbed **Little hierarchy problem** [10]: if we play at adding generic contributions from New Physics at a scale Λ_N in the form of higher dimensional operators, they will necessarily induce corrections to well measured quantities in the SM, like the ρ parameter. All in all, the precision constraints would require $\Lambda_N > 5 \div 10$ TeV: why is this scale one order of magnitude larger than the required cutoff in the BEH mass term? The above considerations are at the root of the naturalness argument, arguably the strongest motivation for the presence of New Physics around the TeV scale.

Why is the scalar mass so sensitive to the cutoff of the theory? The ultimate reason is related so symmetries! Did you know that the responsible is the spin, i.e. rotation symmetry? In fact, for particles with non-zero spin, like spin-1 vector bosons and spin-$\frac{1}{2}$ fermions, the mass is protected against quadratically-divergent corrections, hence no hierarchy problem arises.

2.2.1 Spin 1

Spin 1 particles have three possible spin configurations. In the case of a massless vector boson, the lagrangian is invariant under gauge symmetries, and one can use a gauge transformation to remove one of the degrees of freedom. Physically, this corresponds to the fact that the photon has only two circular polarisations, but no longitudinal modes. Adding a non-zero mass to the vector, gauge symmetries are broken and the vector regains its lost polarisation. Thus there is a fundamental difference between massless and massive vector bosons.

Even for massive vectors, loop corrections to the masses are protected. In fact, for large momenta running in the loop (near the cutoff), the mass of the vector can be neglected, thus a gauge symmetry is restored. The restored gauge symmetry makes the loop correction vanish near the cutoff.

2.2.2 Spin 1/2

Spin 1/2 particles have two spin configurations. For fermions, a mass term connects left-handed to right-handed chiralities. For massless fermions, the two polarisations are independent, thus they can be considered as physically distinct fields. This fact increases the number of symmetries of the system!

Again, in the limit of small mass, the two polarisation decouple and the loop correction must vanish!

2.2.3 Spin 0

No such argument applies to scalar fields, which have a single spin configuration!

2.3 Symmetries and New Physics

The naturalness argument requires the presence of a symmetry (protection mechanism) to shield the Higgs mass from quadratically divergent loop corrections. We have seen that masses can be protected by the presence of spin, or more specifically by chirality in the case of fermions and gauge symmetries in the case of vectors. Ultimately, every new physics scenario tries to apply these two symmetries on the Higgs boson!

- **Spin-0 related to spin-1/2**: is there a symmetry that relates a scalar to a spin-1/2 particle? If so, the chirality will protect the fermion mass, and our new symmetry will project the protection on the scalar!
  - (a) *Supersymmetry*: it extends the Poincaré algebra to include a symmetry between particles with different spins.
- **Replace a spin-0 with spin-1/2**: can we trade the scalar Higgs for fermions? QCD does it: it is a theory based on quarks and vector gluons, however, due to the strong interacting regime, fermions form mesons which are scalars.
  - (b) *Technicolour/Composite Higgs models* [15, 34]: the Higgs emerges as a composite state (meson) of new fermions bound together by a new strong interaction.
- **Replace a spin-0 with a spin-1 (vector)**: can we use gauge symmetries to protect the Higgs, i.e. embed the BEH scalar into a gauge boson?
  - (c) *Extra dimensions* [8]: in models with extra spacial dimensions, a vector boson has more then three polarisations. The extra polarisations appear, from the 4-dimensional point of view, as scalars, however they are secretly part of a vector and thus protected by extra dimensional gauge symmetries.

– **A special spin-0, global symmetries**: can global symmetries have any role in protecting the Higgs boson?

(d) *Little Higgs models* [42]: the Higgs arises as the Goldstone boson of a spontaneously broken global symmetry. The global symmetry is not exact, thus the Higgs develops a mass, however the model can be engineered to have finite one-loop corrections. It does not work beyond one loop!

3 Supersymmetry: A Fermion-Boson Symmetry

Supersymmetry is a symmetry that relates fermions and bosons to each other. It is useful to address the naturalness problem because it can associate the BEH scalar with a fermion: as the two partners share the same physical properties, the chiral symmetry which protects the fermion mass will also protect the scalar partner mass.

Let's consider a spin-0 quantum state $|s>$: supersymmetry can be thought of in terms of an operator Q, which transforms the scalar state into a spin-1/2 state $|f>$:

$$Q|s>=|f> \ . \tag{25}$$

In order for the equation to respect rotational invariance, the operator Q must carry spin-1/2, thus transforming as a spinor under Lorentz transformations: it is a fermionic operator. The minimal spin-1/2 representation of the Lorentz group is a Weyl fermion, i.e. a 2-component chiral fermion. One can however construct supersymmetric theories with any number of chiral generators Q^i (extended supersymmetry). Q is a fermion, therefore it will respect anti-commutation relations:

$$\{Q, \bar{Q}\} = -2\sigma^\mu p_\mu \,, \quad \{Q, Q\} = 0 \,, \quad \{\bar{Q}, \bar{Q}\} = 0 \,. \tag{26}$$

Furthermore, being a spin-1/2 object, it has the following commutation properties with the position and momentum operators:

$$[Q, p_\mu] = 0 \,, \qquad [\bar{Q}, p_\mu] = 0 \,. \tag{27}$$

So far, we have only used the transformation properties of a spinor under space-time symmetries. What we obtained is that we can define a closed algebra including the usual Poincaré algebra, extended by the addition of the fermionic operator Q. In this sense supersymmetry is an extension of space-time symmetries.

3.1 How to Construct a Supersymmetric Quantum Field Theory?

The most straightforward way would be to write down a theory containing a scalar field φ and a chiral fermion χ, corresponding to the two related quantum states, and

then derive the transformation properties of the fields under the operator Q. This procedure, however, turns out to be quite lengthy.

A shortcut is offered by the previous observation that the operator Q can be formally included into the Poincaré algebra: this fact leads to the introduction of *superfields* [35, 44]. Like space-time co-ordinates x^μ correspond to the momentum operators p^μ, we can think of associating co-ordinates θ (and $\bar{\theta}$) to the supersymmetric operators Q (and its conjugate $\bar{Q}$). What we are doing, therefore, is to enlarge space-time by adding some "funny" new co-ordinates: in fact, Q being a spinor implies that θ and $\bar{\theta}$ must be anti-commuting co-ordinates (or Grassmann variables). In some sense, they are spinors themselves, thus they carry an index labelling its two components (in the following we will often omit the spinor indices, assuming that they are always properly contracted). A field living in the superspace $\{x^\mu, \theta\}$, a superfield, is therefore simply a function of the extended set of co-ordinates x^μ, θ and $\bar{\theta}$. For a "scalar" superfield:

$$S(x^\mu, \theta, \bar{\theta}) \,. \tag{28}$$

One can also define functions that transform non-trivially under Lorentz transformations: for instance, a "spinor" superfield W^α or a "vector" superfield W^μ, etc. Note that the Lorentz properties of the superfields do not have anything to do with the spin of the fields they encode! The next step would be to understand how a superfield transforms under supersymmetric transformations generated by Q.

Now, θ is a special kind of co-ordinate because it anti-commutes with itself and has two independent components: therefore, powers of θ^n with $n > 2$ must vanish. This is due to the fact that when contracting more than two identical spinors, at least two of them must have their spin aligned and such configuration is forbidden by Pauli's exclusion principle. This means that any superfield can be expanded in a finite series in powers of the super-coordinate θ. The most general expansion for our "scalar" superfield reads:

$$\begin{aligned} S(x^\mu, \theta, \bar{\theta}) = a + \theta\xi + \bar{\theta}\bar{\chi} + \theta\theta b + \bar{\theta}\bar{\theta}c + \\ \bar{\theta}\bar{\sigma}^\mu\theta v_\mu + \bar{\theta}\bar{\theta}\theta\eta + \theta\theta\bar{\theta}\bar{\zeta} + \theta\theta\bar{\theta}\bar{\theta}d \,, \end{aligned} \tag{29}$$

where σ^μ are, as usual, the Pauli matrices. The coefficients of the expansion are standard fields, functions of x^μ only: a, b, c, and d are scalar (spin-0) fields; ξ, χ, η and ζ are chiral fermions and v_μ is a vector. However, one should define some more minimal representations of the supersymmetric algebra, i.e. superfields that have less independent components than the general expansion. This selection is similar to the definition of spins: even though a spinor in 4 dimensions has 4 components, the minimal representation is a 2-component (chiral) Weyl fermion. The minimal superfield is the *chiral superfield* Φ, defined as

$$\Phi(y^\mu, \theta) = \varphi(y^\mu) + \sqrt{2}\,\theta\chi(y^\mu) + \theta\theta F(y^\mu) \,, \tag{30}$$

where $y^\mu = x^\mu + i\bar{\theta}\bar{\sigma}^\mu\theta$. Note that Φ only depends on $\bar{\theta}$ implicitly via y^μ. The definition of Φ can be formally extracted from the supersymmetric transformation properties of superfields, however the formal treatment is beyond the scope of these lectures. Note also that Φ contains a scalar field φ, a 2-components spinor χ, and an extra field F, whose function will be clear shortly.

The next step is to write an action for the superfield: in addition to the integral over the usual space-time, we need to integrate over the super-coordinate θ. There are, in this case, two possible ways of integrating:

$$\int d^2\theta\,, \qquad \text{and} \qquad \int d^2\theta d^2\bar{\theta}\,, \tag{31}$$

where $\bar{\theta}$ is the hermitian conjugate of θ. Another consequence of the fermionic nature of θ is that the only non-vanishing integrals are:

$$\int d^2\theta\ \theta\theta = 1\,, \quad \int d^2\theta d^2\bar{\theta}\ \theta\theta\bar{\theta}\bar{\theta} = 1\,, \tag{32}$$

where the two integrals are normalised to 1 for convenience. As we can always expand any function of θ and $\bar{\theta}$ in a finite series of terms, the integral definition is such that $\int d^2\theta$ selects the term of the expansion proportional to $\theta\theta$, i.e.

$$\int d^2\theta\ S(x^\mu, \theta, \bar{\theta}) = b + \bar{\theta}\bar{\zeta} + \bar{\theta}\bar{\theta}d\,, \tag{33}$$

while

$$\int d^2\theta\, d^2\bar{\theta}\ S(x^\mu, \theta, \bar{\theta}) = d\,. \tag{34}$$

Here we chose to integrate over a "scalar" superfield because we want the Action to be invariant under Lorentz transformations.

Finally, we need to define a supersymmetric action, which contains an integration over the super-coordinates and is invariant under supersymmetric transformations. There are two possibilities, and they are both important in the definition of supersymmetric theories. On one hand, one can integrate over the whole superspace any "scalar" superfield (note that products of superfields are also superfields, being functions of θ and $\bar{\theta}$):

$$\mathscr{S}_1 = \int d^4x \int d^2\theta d^2\bar{\theta}\ S(x^\mu, \theta, \bar{\theta})\,. \tag{35}$$

As already noted, this selects the $\theta\theta\bar{\theta}\bar{\theta}$ term of the expansion (D-term): you can check that under supersymmetric transformations, such term only picks up total derivatives, which vanish once integrated over the whole space-time. This is enough

to prove that the definition $\mathscr{S}_1$ is a good supersymmetric Action. What are the dimensions in mass of such a term? From the expansion in Eq. (29), as scalar fields like a have dimension 1 and fermion fields like ξ have dimension $3/2$, we can deduce that θ has dimension $-1/2$. Following our normalisation of the integrals, $d\theta^2$ must have dimension 1. The action is a pure number, therefore the superfield (or combination of superfields) we integrate over in $\mathscr{S}_1$ must have dimension 2 in mass: this fact will play an important role in defining a renormalisable supersymmetric action.

The second possibility is to integrate over $d^2\theta$ a chiral superfield after setting $\bar{\theta} = 0$, i.e.

$$\mathscr{S}_2 = \int d^4x \int d^2\theta \, \Phi(x^\mu, \theta) + h.c.\,. \tag{36}$$

This selects the F-term in the expansion in Eq. (30), which can be shown to also transform with a total derivative under supersymmetric transformations. Counting dimensions, the chiral superfield Φ (or product of chiral superfields) must have dimension in mass 3 for the action $\mathscr{S}_2$ to be a pure number.

3.1.1 Supersymmetric Action for a Chiral Superfield

Let's consider a chiral superfield Φ, and try to build an Action for its field components. The full expansion of Φ reads

$$\begin{aligned}\Phi(y^\mu, \theta) = \varphi + \sqrt{2}\theta\chi + \theta\theta F - i\,\bar{\theta}\bar{\sigma}^\mu\theta\partial_\mu\varphi + \\ \frac{i}{\sqrt{2}}\theta\theta\bar{\theta}\bar{\sigma}^\mu\partial_\mu\chi - \frac{1}{4}\theta\theta\bar{\theta}\bar{\theta}\partial^\mu\partial_\mu\varphi\,,\end{aligned} \tag{37}$$

where all the fields φ, χ and F are intended to be functions of x^μ, and the $\bar{\theta}$ dependence is explicit.

Let's first use $\mathscr{S}_1$: the first attempt would be to integrate over a single chiral superfield

$$\int d^4x \int d^2\theta d^2\bar{\theta}\, \Phi(y^\mu, \theta) = -\int d^4x\, \frac{1}{4}\partial^\mu\partial_\mu\varphi \;= 0\,, \tag{38}$$

because the $\theta\theta\bar{\theta}\bar{\theta}$ term is a total derivative. The same would be true for any chiral superfield, including products of two or more superfields. As a second try, we can integrate over the product of a chiral superfield with its complex conjugate (which is not a chiral superfield):

$$\int d^4x \int d^2\theta d^2\bar{\theta}\, \Phi^\dagger\Phi = \int d^4x\, (\partial^\mu\varphi)^\dagger\partial_\mu\varphi - i\,\bar{\chi}\bar{\sigma}^\mu\partial_\mu\chi + F^*F\,. \tag{39}$$

This looks like the kinetic term for a scalar field φ and a fermion χ. The extra field F does not have any derivative, thus it is not a dynamic field (auxiliary field), and it can be easily integrated out. Note also that this is the lowest dimensional object we can write down, and the only one which is renormalisable, because it saturates the dimension of the integrand (2) and therefore its coefficient is a pure number. Adding the product of more than two superfields would induce higher dimensional operators, because we would need to divide by a mass scale to compensate for the dimension in mass of the extra superfields. As a summary, we discovered that the action $\mathscr{S}_1$ can be used to define kinetic terms for the field coefficients of a chiral superfields, plus a static F field (auxiliary field) which can be integrated out.

To use $\mathscr{S}_2$, we need chiral superfields: after setting $\bar{\theta} = 0$, their expansion simplifies

$$\Phi(x^\mu, \theta) = \varphi + \sqrt{2}\,\theta\chi + \theta\theta F\,. \tag{40}$$

The most general action will therefore be

$$\int d^4x \int d^2\theta \left[\frac{1}{2}\mu\ \Phi\Phi + \frac{1}{3}y\ \Phi\Phi\Phi\right], \tag{41}$$

where we have kept only the normalisable interaction (and neglected a linear term). Here μ has dimension of mass, while y is a pure number. The integral selects the terms in the expansion proportional to $\theta\theta$. There are two possibilities: either we take an F component from one superfield and the scalar ones from the remaining ones, or we select a fermion from two superfields, and scalars from the remaining ones:

$$\begin{aligned}\int d^4x \int d^2\theta \left[\tfrac{1}{2}\mu\ \Phi\Phi + \tfrac{1}{3}y\ \Phi\Phi\Phi + h.c.\right] \\ = \int d^4x \left[F\left(\mu\varphi + y\varphi^2\right) - \left(\tfrac{1}{2}\mu + y\varphi\right)\chi\chi + h.c.\right].\end{aligned} \tag{42}$$

This action term generates a mass for the fermion, an interaction of fermions with the scalar, and an interactions of the F term with scalars. The F terms is not dynamical, as it does not have derivatives, therefore we can use the equations of motion to calculate it in terms of the other fields, and eliminate it from the action. Putting together this term with the kinetic terms in Eq. (39), the equation of motion for the auxiliary field F reads

$$F^* + \left(\mu\varphi + y\varphi^2\right) = 0\,. \tag{43}$$

Therefore

$$F^*F + F\left(\mu\varphi + y\varphi^2\right) + F^*\left(\mu\varphi + y\varphi^2\right)^* = -\left|\mu\varphi + y\varphi^2\right|^2 = -F^*F\,; \tag{44}$$

where the first term comes from the kinetic action, and the last two from the superpotential action in Eq. (42) and its complex conjugate. Putting all the pieces together, we obtain the following action for the scalar φ and the fermion χ fields:

$$\mathscr{S}_\Phi = \int d^4x \left[(\partial^\mu \varphi)^\dagger \partial_\mu \varphi - i\, \bar{\chi} \bar{\sigma}^\mu \partial_\mu \chi \right.$$
$$\left. - \left|\mu\varphi + y\varphi^2\right|^2 - \left(\frac{1}{2}\mu + y\varphi\right) \chi\chi + h.c. \right]. \tag{45}$$

Note that both the complex scalar φ and the chiral fermion have mass $m_\varphi = m_\chi = \mu$: this is the first consequence of supersymmetry.

3.1.2 General Expressions

The derivation above can be extended to a generic number of superfields. If we have N chiral superfields Φ_i, the most general action can be written as

$$\mathscr{S} = \int d^4x \left\{ \int d^2\theta d^2\,\bar{\theta}\, \Phi_i^\dagger \Phi_i + \int d^2\theta\; W(\Phi_i) + h.c. \right\}, \tag{46}$$

where the *superpotential* W is a function of the superfields (if renormalisable, it can contain up to trilinear terms). Besides the kinetic terms for scalars and fermions, the action will contain the following interactions:

$$\mathscr{S}_W = \int d^4x \left[-\sum_i \left| \frac{\partial W(\varphi)}{\partial \varphi_i} \right|^2 - \frac{1}{2} \sum_{i,j} \frac{\partial^2 W(\varphi)}{\partial \varphi_i \partial \varphi_j} \chi^i \chi^j + h.c. \right], \tag{47}$$

where the first term (the scalar potential) comes from the integration of the F terms.

4 Minimal Supersymmetric Standard Model

Now that we are familiar with the construction of supersymmetric actions with the use of superfields, we can attempt to construct a supersymmetric action that contains the Standard Model one. A chiral superfield contains a Weyl spinor and a complex scalar: it is therefore an ideal candidate to describe the SM fermions and their superpartners. Recall that the Standard Model is indeed build in terms of Weyl spinors (chiral fermions), because left- and right-handed components of the same fermion transform differently under the electro-weak gauge symmetry.

Vector bosons, on the other hand, cannot be expressed in terms of chiral superfields. One way to deal with them is to define another close representation of the supersymmetric algebra: the *real superfield* $V(x^\mu, \theta, \bar{\theta})$ which is a general "scalar" superfield with the constraint

$$V = V^\dagger .$$

The superfield V is often called *vector superfield* for the simple reason that the vector coefficient v^μ in the expansion (29) is not projected out. However, we are still using a "scalar" superfield to construct it! One can use the same actions $\mathscr{S}_1$ and $\mathscr{S}_2$ to define a proper action for the components of the real superfield (which turn out to be a vector v^μ and a chiral fermion λ), and use them to implement gauge interactions. This construction will be left out of this series of lectures, and we refer the reader to Ref. [35, 44] for more details.

4.1 Supersymmetry in Action: Naturalness of the Top Loop

Let's start from the problem that motivated the introduction of supersymmetry in the first place, i.e. the naturalness of the Higgs mass which is endangered, in the SM, by the divergent loop contribution of the top Yukawa coupling (among others). We will first consider the supersymmetric version of the top Yukawa:

$$\mathscr{S}_{top} = \int d^4x \, \left[-y_t \chi_Q \varphi_H \chi_{t_R} + h.c. \right] , \tag{48}$$

where we have re-written the interaction in Eq. (21) in terms of Weyl spinors: χ_Q for the left-handed doublet and χ_{t_R} for the right-handed singlet. The BEH field is called here φ_H to distinguish it from the scalar super-partners of the tops. A supersymmetric version must contain three chiral superfields which have the same quantum numbers as the SM fields:

$$\Phi_H = \varphi_H + \sqrt{2}\theta\chi_H + \theta\theta F_H , \tag{49}$$

$$\Phi_Q = \varphi_Q + \sqrt{2}\theta\chi_Q + \theta\theta F_Q , \tag{50}$$

$$\Phi_{t_R} = \varphi_{t_R} + \sqrt{2}\theta\chi_{t_R} + \theta\theta F_{t_R} . \tag{51}$$

The most general superpotential one can write down is:

$$W = y_t \Phi_Q \Phi_H \Phi_{t_R} . \tag{52}$$

This is the only term, compatible with the quantum numbers (colour, hypercharge and SU(2) properties) of the superfields, and being renormalisable. Using the derivation in Eq. (47), the supersymmetric interactions generated by such a term are:

$$\begin{aligned} \mathscr{S}_{susy-top} = \int d^4x \, \Big[& - y_t \left(\varphi_H \chi_Q \chi_{t_R} + \varphi_Q \chi_H \chi_{t_R} + \varphi_{t_R} \chi_Q \chi_H + h.c. \right) \\ & - y_t^2 \left(\varphi_Q^* \varphi_H^* \, \varphi_Q \varphi_H + \varphi_H^* \varphi_{t_R}^* \, \varphi_H \varphi_{t_R} + \varphi_Q^* \varphi_{t_R}^* \, \varphi_Q \varphi_{t_R} \right) \Big] . \end{aligned} \tag{53}$$

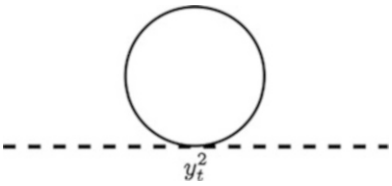

Fig. 2 Stop loop contribution to the Higgs mass

The first term is exactly the SM top Yukawa coupling, thus this superpotential is a proper supersymmetric extension of the top Yukawa sector. The BEH field φ_H has two additional 4-scalar interactions with φ_Q and φ_{t_R}. Such interactions, will contribute to the loop corrections to the Higgs mass, see Fig. 2. Each loop will contribute (here we assign a mass m to the scalar top partners):

$$\begin{aligned} -i\delta m_\phi^2 &= -3iy_t^2 \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2} \\ &= \frac{3y_t^2}{16\pi^4} 2\pi^2 i \int k_E^3 dk_E \frac{1}{-k_E^2 - m^2} \\ &= -\frac{3y_t^2}{16\pi^2} \int dk_E^2 \frac{k_E^2}{k_E^2 + m^2} \\ &= -\frac{3y_t^2}{16\pi^2} \int_0^{\Lambda^2} dk_E^2 \left(1 - \frac{m^2}{k_E^2 + m^2}\right) \\ &= -\frac{3y_t^2}{16\pi^2} \left(\Lambda^2 - m^2 \log \frac{\Lambda^2 + m^2}{m^2}\right) . \end{aligned} \tag{54}$$

Summing the contribution of the two scalar tops (stops) with the top one in Eq. (22), the quadratically divergent term cancels out, and we are left with log divergent terms

$$\delta m_\phi^2 = -\frac{3y_t^2}{16\pi^2} \left(m_Q^2 \log \frac{\Lambda^2}{m_Q^2} + m_{t_R}^2 \log \frac{\Lambda^2}{m_{t_R}^2}\right) . \tag{55}$$

Note here that the two masses m_Q and m_{t_R} we arbitrarily assigned to the stops are not supersymmetric: in fact, the fermionic top is still massless because its mass can only be generated by the EWSB. On the other hand, masses for the scalar tops are allowed by gauge invariance. We will comment on the origin and significance of such terms later on: for now notice that in the case of exact supersymmetry (i.e. $m_Q = m_{t_R} = 0$) the loop correction to the BEH field mass would be exactly zero.

Table 1 Chiral superfield content of the MSSM. For completeness, we also add their Baryon and Lepton numbers B and L

	Label	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$	3B	L	$(-1)^{3(B-L)}$
r-h electron	e_R	**1**	**1**	1	0	-1	$-$
l-h leptons	L	**1**	**2**	$-1/2$	0	1	$-$
r-h up quark	u_R	**3**	**1**	$-2/3$	-1	0	$-$
r-h down quark	d_R	**3**	**1**	1/3	-1	0	$-$
l-h quarks	Q	**3**	**2**	1/6	1	0	$-$
Higgs (up)	H_u	**1**	**2**	1/2	0	0	$+$
Higgs (down)	H_d	**1**	**2**	$-1/2$	0	0	$+$

4.2 The MSSM

In a minimal supersymmetric version of the Standard Model [35], besides supersymmetric gauge interactions that we have not described here, we need to promote each SM field to a chiral superfield (listed in Table 1). Note the presence of two Higgs doublets, H_u and H_d with opposite hypercharges. If we had only one Higgs, say H_u, we would only be able to write Yukawa interactions for up quarks:

$$W_u = y_u \Phi_{H_u} \Phi_Q \Phi_{u_R} \,. \tag{56}$$

The other Yukawas in the SM should be written in terms of $\varphi^*_{H_u}$, which is contained in H^*_u (which is not a chiral superfield!). Thus, in order to preserve supersymmetry, we are forced to introduce a second Higgs doublet, with opposite sign hypercharge that will couple to down-type fermions:

$$W_d = y_d \Phi_{H_d} \Phi_Q \Phi_{d_R} + y_e \Phi_{H_d} \Phi_L \Phi_{e_R} \,. \tag{57}$$

We can also add a bilinear in the two Higgs superfields:

$$W_H = \mu \Phi_{H_u} \Phi_{H_d} \,, \tag{58}$$

which will generate a mass for the two Higgs scalars $m_{H_u} = m_{H_d} = \mu$.

There is another reason why two Higgses are needed: in the SM, a complete generation of fermions is anomaly free. In supersymmetry, the Higgs superfield will contain a new fermion doublet, the superpartner of the Higgs. The presence of a single higgsino would generate anomalies: the role of the second Higgs is therefore to cancel such anomalies. The three terms listed here should be the only superpotential terms, because they generate the needed fermion yukawa couplings and a mass for the Higgses. However this is not the case generically.

4.2.1 Troubleshooting 1: Unwanted Superpotential Terms (the Need for R-Parity, and Dark Matter)

In addition to the Yukawa interactions, the superfield content in Table 1 allows for many more "dangerous" terms to be added. The game here is to add all combinations of two or three superfields that can lead to a term invariant under all gauge symmetries. For instance, one can add an operator made of three quark singlets:

$$\Phi_{u_R}\Phi_{d_R}\Phi_{d_R}\,; \tag{59}$$

an operator of this kind would be forbidden in the SM because it contains three fermions.

Also, Φ_L and Φ_{H_d} have exactly the same quantum numbers, thus they can be interchanged:

$$\Phi_{H_d}\Phi_{H_d}\Phi_{e_R}\,, \quad \ldots \tag{60}$$

What symmetries of the SM are violated by such superpotential terms?

The first kind will violate Baryon number (the operator has a net baryon number -1, like an anti-neutron); the second kind violates lepton number. Such terms are very dangerous because, among other things, they can mediate the proton decay. Recall that both Baryon and Lepton number conservation are an accidental consequence of the matter content of the SM. In supersymmetric extensions of the SM, such an accident does not occur. The basic reason for this is that in the MSSM, both the fermions and the Higgs boson are traded with "scalar" superfields, therefore Baryon and Lepton numbers, which are a relic of the flavour symmetry in the SM, are completely removed.

At this point there are two options: one may either assume that the "unwanted" terms are small for some unknown reason, in order to satisfy the bounds. The other route is to add a symmetry that would forbid such terms: this symmetry necessarily is a shadow of Baryon and Lepton symmetries, and goes under the name of **R-parity**.

4.2.2 R-Parity and Dark Matter

The solution is to somehow impose Baryon and Lepton number conservation by hand. We therefore impose a Z_2 parity on the superfields, defined as:

$$P_M = (-1)^{3(B-L)}\,; \tag{61}$$

it is called *matter parity*, and it is defined in terms of $B-L$ because this combination is anomaly free in the SM and thus it is guaranteed to be conserved. Requiring the superpotential to be even under matter parity eliminates all the unwanted terms.

We can further elaborate this symmetry: any action term must contain an even number of fermions, so we can redefine matter parity by adding an extra "-1" for fermions, without affecting the allowed interaction terms:

$$P_R = (-1)^{3(B-L)+2s}\,, \tag{62}$$

where s is the spin of the field. This parity now acts on the field components of the superfields (thus it is not compatible with supersymmetry), and it is called *R-parity*. Note that now scalars and bosons in the same superfield have opposite R-parity; furthermore, all SM states (matter fermions and scalar Higgses) have R-parity $+1$, while the supersymetric partners (squarks, sleptons and higgsinos) have R-parity -1. This implies that the lightest supersymmetric particle is stable, because it cannot decay into SM states only!

Can it play the role of Dark Matter [11]?

A few comments are in order here: as we have seen, R-parity is needed to save the MSSM from fast proton decays, so it is added by hand on the supersymmetric action. While the presence of a stable Dark Matter candidate is a consequence of imposing R-parity, it should not be considered as a prediction of supersymmetry itself. In fact, the most general supersymmetric MSSM would not contain any dark matter candidate! Supersymmetry also contains an internal symmetry, called R-symmetry, which is a phase transformation of the θ co-ordinates. It should be stressed here that this R-symmetry has nothing to do with R-parity, but it is simply related to the fact that two fermions are always needed to write a Lorentz invariant interaction term.

Note also that R-parity is not exactly equivalent to imposing Baryon and Lepton conservation: in fact, it is a parity, therefore only terms with an odd Baryon or Lepton number are forbidden. For instance, a Majorana mass for neutrinos, which would violate lepton number by two units, is allowed. Thus the typical neutrino mass mechanisms (small Yukawa couplings with a singlet, see-saw, etc.) can be implemented in the MSSM.

4.2.3 Troubleshooting 2: Supersymmetry Cannot Be Exact! Where Do We Expect the Superpartners?

Another problem is that supersymmetry is not an exact symmetry, because it would predict that SM states and their partners have the same mass (we are pretty confident that there are no scalar electrons around!).

One way to break supersymmetry without spoiling its nice properties (mainly the cancellation of divergences), is to add only "mass terms", i.e. couplings with a positive mass dimension: the reason behind is that at high energies, well above the supersymmetry breaking mass scales, supersymmetry is restored, thus the divergences still cancel out! This principle is called *soft supersymmetry breaking*. We should also be careful not to violate R-parity if we want a Dark Matter candidate in the model. The allowed terms are therefore:

- Higgs mass terms: $-m^2_{H_u}\varphi^*_{H_u}\varphi_{H_u} - m^2_{H_d}\varphi^*_{H_d}\varphi_{H_d}$;
- Scalar quark and lepton masses: $-\tilde{m}^2_Q\varphi^*_Q\varphi_Q - \tilde{m}^2_{t_R}\varphi^*_{t_R}\varphi_{t_R} + \ldots$;
- Trilinear scalar couplings (in the same form as Yukawa couplings): $A\varphi_{H_u}\varphi_Q\varphi_{t_R} + \ldots$;
- Gaugino masses (masses for the fermion partners of gauge bosons).

Note that a huge number of soft supersymmetry breaking terms can be added to the MSSM (more that 120!). In order to study the phenomenology, one needs to make simplifying assumptions or develop a mechanism of supersymmetry breaking.

We will not go into the details of the supersymmetry breaking mechanisms, however a few comments are in order:

- If supersymmetry is spontaneously broken, then the Goldstone theorem (extended to fermionic symmetries) would predict the presence of a massless fermionic field (called *goldstino*). Such object presents a challenge for the phenomenology: one way out is to promote supersymmetry to a local symmetry (*supergravity*), where the goldstino is eaten by the spin-3/2 partner of the graviton to give rise to a massive spin-3/2 state (*gravitino*). In some models, it is the gravitino that plays the role of the Dark Matter, being the lightest supersymmetric (thus R-parity odd) state.
- The most popular supersymmetry breaking mechanisms are *m-SUGRA*, *AMSB* and *GMSB*: m-SUGRA is based on a naive assumption that supergravity would couple equally to all states of a given spin, therefore if gravity breaks supersymmetry then all particles with same spin would receive the same supersymmetry breaking mass. In AMSB (Anomaly Mediated Supersymmetry Breaking), it is gravitational anomalies that break supersymmetry: this is therefore a realistic model of breaking mediated by gravitational interactions. In GMSB (Gauge Mediated SB), the basic assumption is that supersymmetry breaking (which may be generated dynamically *á la* Technicolour) is mediated to the MSSM via gauge interactions. Such models have very different predictions on the low energy spectrum of the MSSM.

One important question, however, independent on the origin of supersymmetry breaking is the value of the scale where the masses of the superpartners should be. The handle we have on this question is the naturalness argument. In fact, the partner masses enter the BEH field mass via loop corrections, like the ones generated by the top and stop loops:

$$m^2_{BEH} \sim \mu^2 - \frac{3y_t^2}{16\pi^2}\left(m^2_Q \log\frac{\Lambda^2}{m^2_Q} + m^2_{t_R}\log\frac{\Lambda^2}{m^2_{t_R}}\right). \tag{63}$$

This formula is very important for the MSSM phenomenology. On one hand, we need the stop loop to give a contribution not far from the measured value of the Higgs mass: this observation would tell us that the masses m_Q and m_{t_R} should not be too far from the TeV scale. Another noteworthy element is the presence of the μ

term, which is a supersymmetric coupling, and which gives a positive mass to the BEH field: in order for the model to feature EWSB, we would need μ to be smaller than the top loop contribution. So the question is: why should a supersymetric coupling, which can take any value between 0 to the Planck mass, know about the scale of supersymmetry breaking? This question is called the μ-problem. One possible solution is to extend the MSSM with a singlet field, and use it to generate a μ-term via its vacuum expectation value [25]: this model is called NMSSM.

5 Extra Dimensions: Warm Up

Extra dimensions are also a simple extension of the space-time symmetries, by extending the usual 4-dimensional space-time with the addition of extra space dimensions. This idea was first proposed at the beginning of last century in an attempt to unify QED with General Relativity. While this project failed, extra dimensions appeared again when string theory was formulated: in order to have a mathematically consistent theory, string theory should be formulated on least 10 dimensions. This dimensions are certainly not part of the space we can normally probe (as you can check with your own eyes), so their effects must be somehow hidden and appear only above a certain energy scale. Initially string theory aimed at a consistent formulation of quantum gravity, therefore the scale where their effect would appear was assumed to be the Planck scale. In the early 1990s, I. Antoniadis and K. Benakli [6,7] realised that extra dimensions may actually appear at much lower scales, possibly accessible to colliders, thus they ignited an active research program aimed at exploring the phenomenological implications of and new mechanisms available in extra dimensional space-time.

The first issue to be addressed is the mechanism to hide the extra dimensions at high scale: the simplest way is to postulate that the extra space is compact, another possibility is to modify the geometry of the space so that extra energy is needed for particles to propagate in the curved background. In the following we will see an example of both.

5.1 A 5D Scalar Field

Before detailing some of the models of BSM based on extra dimensions, we need to understand how their effect would appear at low energy. We can start our exploration with a scalar (spin-0) field with a single extra space co-ordinate: the action is simply extended to

$$\mathscr{S}_s = \int d^5x \; (\partial^M \Phi)^\dagger \partial_M \Phi - M^2 \Phi^\dagger \Phi \,, \tag{64}$$

where $M = \mu, 5$ labels the 5 directions in space-time (the derivative is promoted to a 5-vector), and

$$\Phi = \Phi(x^\mu, x_5)\,. \tag{65}$$

Here we are also assuming that the metric on the 5D space is an extension of Minkowski (flat space). From the above action we can derive the usual Klein-Gordon equation of motion:

$$-\partial^M \partial_M \Phi - M^2 \Phi = -\partial^\mu \partial_\mu \Phi + \partial_5^2 \Phi - M^2 \Phi = 0\,. \tag{66}$$

The simplest compact space is a circle, i.e. a space where we impose periodic conditions on the fields:

$$\Phi(x^\mu, x_5 + 2\pi R) = e^{i\alpha_\Phi} \Phi(x^\mu, x_5)\,; \tag{67}$$

in general a non-zero phase α_Φ (Scherk-Schwarz phase) may be imposed, for simplicity here we will only consider periodic fields and we will set $\alpha_\Phi = 0$. If we want to go to momentum space, along the visible directions the usual Fourier transform applies; on the other hand, along x_5 we need to Fourier expand in a series of functions (the domain of the function of x_5 is finite!). Therefore, one can rewrite the scalar field as

$$\Phi(x^\mu, x_5) = \int \frac{d^4 p}{(2\pi)^4}\, e^{i p_\mu x^\mu} \sum_n f_n(x_5) \varphi_n(p^\mu)\,; \tag{68}$$

where p^μ is the usual 4D momentum, f_n is a complete set of functions on the compact extra space (*wave functions*), and the "coefficients" $\varphi_n(p^\mu)$ can be interpreted as 4D fields (*Kaluza Klein modes*). Plugging this expansion in the equation of motion, we obtain a set of equations for f_n:

$$(p^2 - M^2) f_n - \partial_5^2 f_n = 0 \tag{69}$$

whose solutions are

$$\sin x_5 \sqrt{p^2 - M^2}\,, \qquad \cos x_5 \sqrt{p^2 - M^2}\,. \tag{70}$$

The periodicity implies that

$$\sqrt{p^2 - M^2} = n/R \tag{71}$$

where n is positive integer. Now p is the usual 4 dimensional momentum, therefore we can interpret p^2 as the 4D mass of the 4D field, and the above equation yields

$$p^2 = m_n^2 = \frac{n^2}{R^2} + M^2 = n^2 m_{KK}^2 + M^2\,, \quad m_{KK} = 1/R\,. \tag{72}$$

The complete expansion of the field is then (where we have properly normalised the wave functions f_n)

$$\Phi(p^\mu, x^5) = \frac{1}{2\pi R}\varphi_0 + \sum_{n=1}^{\infty} \frac{\cos n x_5/R}{\pi R}\varphi_{n,c} + \sum_{n=1}^{\infty} \frac{\sin n x_5/R}{\pi R}\varphi_{n,s}\,, \tag{73}$$

with effective 4D action

$$\mathscr{S}_s = \int d^4x \left[(\partial^\mu \varphi_0)^\dagger \partial_\mu \varphi_0 + \right.$$
$$\left. + \sum_n (\partial^\mu \varphi_{n,c/s})^\dagger \partial_\mu \varphi_{n,c/s} - (M^2 + n^2 m_{KK}^2)\varphi_{n,c/s}^\dagger \varphi_{n,c/s}\right]. \tag{74}$$

5.2 Orbifold

Starting from the circle, more spaces can be defined by using the symmetries of the circle itself: one can in fact identify points mapped one into the other by such symmetry. For instance, the circle is invariant under a mirror symmetry with respect to any diameter: $x_5 \to -x_5$. If a circle is defined for $x_5 \subset [-\pi R, \pi R)$, then the mirror symmetry identifies positive and negative points. The resulting space (the interval) is defined on $x_5 \subset [0, \pi R]$.

On the fields, the orbifold projection means that each field must satisfy:

$$\Phi(p^\mu, -x_5) = \pm\Phi(p^\mu, x_5)\,. \tag{75}$$

Each field is characterised by a sign choice; the wave functions that do not respect the transformation properties are then removed.

$$\Phi^+ = \frac{1}{2\pi R}\varphi_0 + \sum_{n=1}^{\infty} \frac{\cos n x_5/R}{\pi R}\varphi_{n,c}\,; \tag{76}$$

$$\Phi^- = \sum_{n=1}^{\infty} \frac{\sin n x_5/R}{\pi R}\varphi_{n,s}\,. \tag{77}$$

Note that the massless $n = 0$ mode is only present for Φ^+; both choices have a tower of massive states with the same mass but different wave functions.

5.3 A 5D Vector (Gauge) Field

The action for a gauge (vector) boson can again be obtained by a simple extension of the 4D case. A vector field must first be generalised to a 5-vector: $A_M = \{A_\mu, A_5\}$. For an abelian gauge group the action is:

$$\begin{aligned}\mathscr{S}_{gauge} &= \int d^5x \left[-\frac{1}{4} F_{MN} F^{MN}\right] \qquad [\text{here} \quad F_{MN} = \partial_M A_N - \partial_M A_N] \\ &= \int d^5x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} F_{\mu 5} F_5^{\mu}\right] \\ &= \int d^5x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} \partial_\mu A_5 \partial^\mu A_5 + \frac{1}{2} \partial_5 A^\mu \partial_5 A^\mu - \partial_\mu A_5 \partial_5 A^\mu\right]. \end{aligned} \tag{78}$$

The $\mu 5$ term generates a mixing between the 4D vector components A_μ and the 4D scalar term A_5: this is similar to the mixing we obtain in the SM between the massive vectors and the Goldstone components of the BEH field. To simplify the equations, we can add a "gauge fixing" term to the action, which is a total derivative that can cancel out the mixing term and decouple the vector and the scalar. The extra dimensional R_ξ gauge fixing term is then (similar to the ones used in the SM):

$$\mathscr{S}_{GF} = \int d^5x \left[-\frac{1}{2\xi} \left(\partial_\mu A^\mu - \xi \partial_5 A^5\right)^2\right]. \tag{79}$$

The "gauge fixed" action now reads:

$$\begin{aligned}\mathscr{S}_{gauge+GF} = \int d^5x \Bigg[&-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2\xi} (\partial_\mu A^\mu)^2 + \frac{1}{2} \partial_5 A^\mu \partial_5 A^\mu \\ &+ \frac{1}{2} \partial_\mu A_5 \partial^\mu A_5 - \frac{\xi}{2} (\partial_5 A^5)^2\Bigg]; \end{aligned} \tag{80}$$

where A_5 features an action similar to a scalar field.

5.3.1 Vector A_μ

From the action in Eq. (80), the equation of motion for the vector part is

$$\partial^\mu F_{\mu\nu} + \frac{1}{\xi} \partial_\nu \partial^\mu A_\mu - \partial_5^2 A_\nu = 0\,. \tag{81}$$

We can Fourier transform and expand the field as before

$$A_\mu(x^\nu, x_5) = \int \frac{d^4p}{(2\pi)^4}\, e^{ip_\alpha x^\alpha} \sum_n f_n(x_5)\, A_\mu^n(p^\nu)\,; \tag{82}$$

and, assuming that the 4D fields A_μ^n satisfy the usual 4D equation of motion in ξ gauge for a massive state,

$$\partial^\mu F_{\mu\nu} + \frac{1}{\xi}\partial_\nu \partial^\mu A_\mu = -p^2 A_\nu \,. \tag{83}$$

we have the following equation for the wave functions

$$(p^2 + \partial_5^2) f_n = 0 \,, \tag{84}$$

which is the same as in the scalar case (but with $M = 0$). The final KK expansion is therefore analogous to the scalar one. The spectrum contains one massless gauge boson (thus in 4D gauge symmetries are respected), and a tower of massive states. Where do the massive states get the longitudinal polarisation, as there is no BEH field here?

5.3.2 Scalar (and the Extra Dimension "BEH" Mechanism)

The equation of motion for the A_5 scalar reads:

$$(\partial_\mu \partial^\mu - \xi \partial_5^2) A_5 = 0 \,, \tag{85}$$

which is similar to the one for a 5D scalar, with the exception of the parameter ξ. After the usual Fourier expansion, the equation for the wave functions is:

$$\left(\frac{p^2}{\xi} + \partial_5^2\right) f_n = 0 \,, \tag{86}$$

thus the expansion is the same as above, except for the substitution $p^2 \to \frac{p^2}{\xi}$.

The masses will therefore be

$$m_n^2 = \xi n^2 m_{KK}^2 \,, \tag{87}$$

which look like the masses of a Goldstone boson in the "BEH" mechanism in ξ-gauge. Note that the only mode whose mass is independent on ξ is the zero mode $n = 0$. The massive states can be decoupled in the limit $\xi \to \infty$ (Unitary gauge): what we learn, therefore, is that the massive modes of the scalar polarisation A_5 are the Goldstone bosons eaten up by the massive vectors! Compact extra dimensions have, therefore, a build-in BEH mechanism that gives mass to the Kaluza-Klein tower of states. This fact has been used to construct "Higgsless" models [19, 20], where the W and Z are identified with Kaluza-Klein states that pick up their mass without the need for a scalar BEH field (and no Higgs boson).

5.3.3 Gauge Invariance

The 5D action is invariant under a generalised gauge transformation:

$$A_M \to A_M + ig\partial_M \alpha(x^\mu, x_5)\,. \tag{88}$$

The local gauge parameter α must satisfy the same properties as the gauge field A_M, thus it also is a periodic function of x_5. We can therefore Fourier expand both A_μ and α, and write down 4D gauge transformations for each KK mode:

$$A^n_\mu \to A^n_\mu + ig\partial_\mu \alpha^n(x^\mu)\,. \tag{89}$$

Naively, we would expect the presence of an infinite number of gauge groups, however, as shown in the mass spectrum, the extra polarisation "spontaneously breaks" the gauge invariance associated with the massive modes; only the 4D gauge invariance of the massless mode is (explicitly) preserved.

Caveat: the Fourier expansion of the gauge transformation properties is a bit naive, one should really consider the gauge transformation on 5D fields!

5.3.4 Orbifold

We can now extend the analysis to orbifolds. As before, the field must be associated with a parity under the orbifold symmetry. However, the parities of the A_μ and A_5 components are related to each other by the fact that they belong to a vector! So if the orbifold symmetry is $x_5 \to -x_5$ (change sign to the 5th component of the vector but not to the other 4), the parity assignment for the 5D vector must be

$$A_\mu(-x_5) = \pm A_\mu(x_5)\,, \qquad A_5(-x_5) = \mp A_5(x_5)\,, \tag{90}$$

in other words their parity must be opposite! Recall that A_M must be a vector, i.e. share the same transformation properties as space-time co-ordinates, because it appears in the covariant derivative: $D_M = \partial_M - igA_M$.

For a $+$ vector, the scalar is $-$: in this case, the vector contains a massless zero mode and massive vectors (with *cos* wave function), while the scalars only contain a tower of Goldstone bosons (with wave function *sin*).

For a $-$ vector, the scalar is $+$: now the vectors only contain a tower of massive states (*sin*), while the scalars contain a physical massless scalar and a tower of Goldstone bosons (*cos*).

Note that the number of massive states always matches, so that the massive "scalars" always provide for the extra polarisation of the massive vectors. Also that for $-$ vectors, the 4D gauge symmetry is broken, as signalled by the absence of a massless vector in the KK expansion! However, a massless scalar is present! The breaking of gauge symmetries is however constrained: the proper definition

is that the transformation of the A_μ under the orbifold projection is: $A_\mu(-x_5) = G(A_\mu(x_5))$, where G is a gauge transformation. Therefore, one can only break a gauge group $\mathscr{G}$ to a subgroup $\mathscr{H}$ of same rank. As a consequence, U(1) symmetries cannot be broken.

5.4 A 5D Fermion

The Dirac Gamma matrices must be generalised to 5D, i.e. we need to define a set of 5 (not 4) anti-commuting matrices. The natural choice is to promote γ^5 to the role of the gamma matrix for the 5th direction. The minimal spinor is now a 4-component one, and it is not possible to define chiral projections. The action is

$$\mathscr{S}_f = \int d^5x \; i\bar{\Psi}\Gamma^M \partial_M \Psi - m\bar{\Psi}\Psi \,, \tag{91}$$

where the 5D fermion can be described in terms of two 2-component Weyl fermions:

$$\Psi = \begin{pmatrix} \chi \\ \bar{\eta} \end{pmatrix} . \tag{92}$$

In terms of Weyl fermions, the action reads

$$\mathscr{S}_f = \int d^5x \; -i\,\bar{\chi}\bar{\sigma}^\mu \partial_\mu \chi - i\,\eta\sigma^\mu \partial_\mu \bar{\eta} - \bar{\chi}\partial_5 \bar{\eta} + \eta \partial_5 \chi + m(\bar{\chi}\bar{\eta} + \eta\chi) \,; \tag{93}$$

from which we can derive the following equations of motion

$$-i\,\bar{\sigma}^\mu \partial_\mu \chi - \partial_5 \bar{\eta} + m\bar{\eta} = 0 \,, \tag{94}$$

$$-i\sigma^\mu \partial_\mu \bar{\eta} + \partial_5 \chi + m\chi = 0 \,. \tag{95}$$

The KK decomposition is in the form

$$\chi = \sum_n g_n(x_5)\chi_n(x^\mu) \,, \quad \bar{\eta} = \sum_n f_n(x_5)\bar{\eta}(x^\mu) \,, \tag{96}$$

where χ_n and $\bar{\eta}_n$ are usual 4D Weyl spinors.

The usual procedure can be followed: we can plug the expansions in the equations of motion, use the 4D equations of motion to replace derivatives with the 4D momenta and combine the two equations. We obtain that both f_n and g_n must satisfy the same equations of motion as a massive scalar field [21].

Note that on a circle, both chiral fields η and χ have a massless mode! In order to have a massless spectrum that corresponds to the SM fermions, we need to remove one or the other in order to have a chiral spectrum!

5.4.1 Orbifold and Chirality

The orbifold symmetry changes sign to x_5: in order for the kinetic term to be invariant, the parities of χ and $\bar{\eta}$ must be opposite. This is clear for the form of the $\chi\partial_5\eta$ terms in the action, which would be odd (and removed by the projection) otherwise. This implies that only one of the two chiralities will have a zero model.

The massive modes of the two chiralities will be combined to form a massive Dirac fermion. The orbifold is thus an essential ingredient for Model Building! Note also that the mass term is forbidden exactly for the same reason.

5.4.2 Odd Mass Terms

Another possibility is to assume that the mass term is odd under the orbifold symmetry: this is not entirely inconsistent, because the fundamental domain of the orbifold is an interval where the mass is uniform. So, let's force the presence of a mass term!

The most obvious problem we encounter is that the mass term would like to couple the two zero modes to form a Dirac fermion of mass m, however one of the two chiralities is removed.

If we remove the η chirality, the equations of motion for the zero mode in Eq. (95) reduce to:

$$\partial_5 g_0 + m g_0 = 0\,, \qquad g_0(x_5) \sim e^{-m x_5}\,. \tag{97}$$

The wave function of a left-handed mode, therefore, is exponentially localised toward the $x_5 = 0$ boundary of the space (for $m > 0$)

For right-handed zero modes

$$-\partial_5 f_0 + m f_0 = 0\,, \qquad f_0(x_5) \sim e^{+m x_5}\,; \tag{98}$$

thus it is localised toward the other boundary.

This trick allows us to localise the massless modes toward one or the other boundary: this feature has been used to build models where the large hierarchies between fermion masses in the SM (the up quarks weights few MeV, while the top fares 175.000 MeV!) is explained in terms of their exponential localisation [27].

6 First Model: Gauge-Higgs Unification in Flat Space

The goal of this class of models is to embed the BEH field in a gauge symmetry, by promoting it to an A_5 component of a bulk gauge boson; in order to have couplings between the Higgs and the electro-weak gauge bosons, the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge bosons and the Higgs should be unified into a single gauge group G. Such a group

must then be broken down to the SM one, therefore we will need to work with an orbifold compactification (this is also required by the chirality of SM fermions!). The group G must fulfil the following requirements:

- G must contain at least $3_{SU(2)} + 1_{U(1)} + 4_H = 8$ generators;
- At the level of zero modes, only $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ must survive, i.e. the orbifold must break $G \to \mathrm{SU}(2) \times \mathrm{U}(1)$;
- Breaking a gauge group corresponds to assign a parity $+$ for the unbroken generators, and $-$ for the broken ones. This must be done in a consistent way, i.e. a gauge boson can be mapped into itself up to a (global) gauge transformation:

$$A_\mu(-x_5) = U A_\mu(x_5) U^\dagger ,$$

  where U is a gauge transformation of G. In particular, this preserves the rank of the original group G and the rank of the preserved gauge group: to obtain the SM, therefore, G must have rank-2 [18];
- For the scalars, at zero mode level, a doublet of $\mathrm{SU}(2)_L$ with non-zero hypercharge should survive.

An attractive possibility is to use $G = \mathrm{SU}(3)$ [12, 43]: it has rank 2, exactly 8 generators and it can be broken to $\mathrm{SU}(2) \times \mathrm{U}(1)$ with

$$U = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} , \tag{99}$$

written in terms of 3×3 matrix generators. The parity assignments of the gauge components will therefore be:

$$\begin{pmatrix} + & + & - \\ + & + & - \\ - & - & + \end{pmatrix} . \tag{100}$$

The 2×2 block corresponds to SU(2) generators, the $+$ in the lower corner to a U(1) generators, finally the 4 components with parity $-$ will provide the BEH field candidate (A_5 has opposite parities from A_μ), as they transform like a doublet under SU(2).

6.1 Spectrum

The spectrum of vector bosons will contain SU(2) gauge bosons $W^\pm$ and W^3, which contain a zero mode and a tower of massive modes; a U(1) gauge boson B with same spectrum as the SU(2) ones; two charged gauge bosons, with the same quantum

numbers as the BEH field, $C^{\pm}$ and D^0, that have no zero mode and just a tower of massive modes. They are embedded in the SU(3) structure as:

$$A_\mu = \begin{pmatrix} \frac{1}{2}W_\mu^3 - \frac{1}{\sqrt{12}}B_\mu & \frac{1}{\sqrt{2}}W_\mu^+ & \frac{1}{\sqrt{2}}C_\mu^+ \\ \frac{1}{\sqrt{2}}W_\mu^- & -\frac{1}{2}W_\mu^3 - \frac{1}{\sqrt{12}}B_\mu & \frac{1}{\sqrt{2}}D_\mu^0 \\ \frac{1}{\sqrt{2}}C_\mu^- & \frac{1}{\sqrt{2}}D_\mu^0 & \frac{2}{\sqrt{12}}B_\mu \end{pmatrix} . \tag{101}$$

The scalar sector will only contain a massless doublet of $SU(2)_L$ (the KK tower of massive scalars are "eaten" by the massive gauge bosons), that will play the role of the BEH field, embedded in SU(3) as:

$$A_5 = \begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{2}}\phi^+ \\ 0 & 0 & \frac{1}{\sqrt{2}}\phi_0 \\ \frac{1}{\sqrt{2}}\phi^- & \frac{1}{\sqrt{2}}\phi_0^* & 0 \end{pmatrix} . \tag{102}$$

At tree level, the BEH field will not have any potential, because its interactions can only come from the gauge boson action:

$$\mathscr{S} = \int d^5x \ \left[-\tfrac{1}{2}\mathrm{Tr}\, F_{MN}F^{MN}\right] ,$$
$$\text{where} \quad F_{MN} = \partial_M A_N - \partial_N A_M + g(A_M A_N - A_N A_M) . \tag{103}$$

No A_5^2 nor A_5^4 terms are present in this action! So, the potential for the Higgs is generated at one loop. We expect it to be **finite**, because the tree level action does not contain a counter-term either for the mass or quartic coupling! Note that this is true at all orders in perturbation theory!

6.2 Potential Issues

- The Higgs field is a gauge boson, so it couples to all particles with strength dictated by the gauge coupling g. What about fermion masses? To obtain masses below m_W, we can use the mass trick to localise the light quarks towards the two boundaries of the space, in order to reduce the overlap to the Higgs.
- How about the top mass? This is a crucial issue, as the localisation can only suppress the couplings with respect to g. One may use gauge group factors to enhance the coupling (as in Ref. [12]), violate Lorentz invariance in the extra space explicitly [36] or curve the space [15, 32].
- How about the Higgs mass? The potential is one-loop generated, so the mass should be rather small. The precise value depends on the details of the model.

6.3 The BEH Potential in Gauge-Higgs Unification

The BEH potential is generated completely at one loop. Only the zero mode will be sensitive to the eventual negative mass, thus the vacuum solution must be independent on the extra co-ordinate x_5. This implies that no tree level mixing with the heavy gauge bosons will be generated! The reason is that modes with different mass have orthogonal wave functions.

6.3.1 The Hosotani Mechanism

Let's assume that the BEH field does develop a vacuum expectation value that breaks $SU(2)_L \times U(1)_Y \to U(1)_{em}$: the vacuum will have the SU(3) embedding

$$\langle A_5 \rangle = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}}\langle \phi_0 \rangle \\ 0 & \frac{1}{\sqrt{2}}\langle \phi_0 \rangle & 0 \end{pmatrix} . \tag{104}$$

For compact spaces allowing for a Scherk-Schwarz phase, it is always possible to find a gauge transformation $\Omega(x_5)$ such that

$$\langle A_5' \rangle = \Omega(x_5)\langle A_5 \rangle \Omega^\dagger(x_5) + \frac{i}{g}\left(\partial_5 \Omega(x_5)\right)\Omega^\dagger(x_5) = 0\,; \tag{105}$$

so that the Higgs VEV disappears from the action. In fact, for

$$\Omega = e^{i\alpha x_5 \lambda_v}\,, \tag{106}$$

where λ_v is the SU(3) generator the vacuum solution is aligned with (in this case λ_6 of SU(3)) and α is a constant, Eq. (105) gives

$$\langle A_5 \rangle + \frac{i}{g}\frac{i\alpha}{2}\begin{pmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,0 \end{pmatrix} = 0\,, \quad \Rightarrow \quad \alpha = \sqrt{2} g \langle \phi_0 \rangle\,. \tag{107}$$

The same transformation must be applied to the gauge vectors:

$$A_\mu' = \Omega(x_5) A_\mu \Omega^\dagger(x_5)\,. \tag{108}$$

What does it change in the theory? The action is invariant, however the periodicity condition on the field A_μ' is different from before:

$$\begin{aligned} A_\mu'(x_5 + 2\pi R) &= \Omega(x_5 + 2\pi R) A_\mu(x_5) \Omega^\dagger(x_5 + 2\pi R) \\ &= \Omega(x_5 + 2\pi R)\Omega^\dagger(x_5) A_\mu'(x_5) \Omega(x_5)\Omega^\dagger(x_5 + 2\pi R)\,, \end{aligned} \tag{109}$$

where

$$\Omega(x_5 + 2\pi R)\Omega^\dagger(x_5) = e^{i\,2\pi R\alpha\lambda_6}\,. \tag{110}$$

The periodicity condition in the new gauge basis, therefore, contains a non-zero Scherk-Schwarz phase for different components of the SU(3) gauge fields. Note that this re-definition can be interpreted as a modification of the boundary conditions only in orbifolds that do allow for Scherk-Schwarz phases, like the one we considered here. The modified periodicity conditions will also affect the spectrum for the massive gauge bosons and will depend on α (which is proportional to the BEH vacuum solution!). For the $W^\pm$ bosons, the new masses will read:

$$m_n^{W^\pm} = \frac{n+\alpha}{R}\,, \qquad m_{W^\pm} = \frac{\alpha}{R}. \tag{111}$$

As expected the "zero mode" has now acquired a mass proportional to α. The numerical value of α will therefore determine the relation between the SM W mass and the KK mass $m_{KK} = 1/R$.

6.3.2 Numerical Results

The calculation of the potential is rather complex. We know the spectrum as a function of the BEH vacuum α, we can use the Weinberg-Coleman potential:

$$V_{eff}(\alpha) = \pm\frac{1}{2}\sum_i \int \frac{d^4p}{(2\pi)^4}\log[p^2 + M_i^2(\alpha)]\,. \tag{112}$$

Some details of the calculation can be found in [8]. Different fields contained in the theory will give different contributions to the potential. As an example, we report here a potential computed in the framework of the model in Ref. [12], based on SU(3): besides the gauge sector, the model contains a set of bulk fermion fields where the top is embedded into, and a set of massive bulk fermions associated with the light generations which are kept massive by use of twisted boundary conditions. An example of numerical potential, as a function of the BEH vacuum α, can be found in Fig. 3: it's interesting to notice that the contribution of the gauge bosons (red/dashed) and of light fermions (green/dot-dashed) have minima at $\alpha \sim \langle\varphi_0\rangle = 0$, while it is the contribution of the top loops (blue/solid) that generates a non trivial vacuum. From the potential we can also calculate the Higgs boson mass, which is proportional to the second derivative of the potential. The results can be found in Ref. [12]: numerically, the Higgs mass is always fairly small (below 150 GeV). Furthermore, the measured value can be obtained for small values of the parameter $\alpha \sim 1/20$, from which we can extract the expected value of the Kaluza-Klein resonance masses: $m_{KK} \sim 20 \cdot m_W \sim 1.6\,\text{TeV}$.

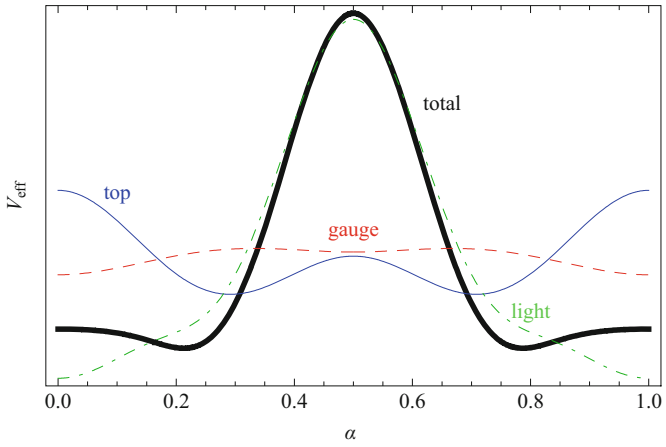

Fig. 3 One-loop Higgs potential for the model in [12]: the *thick line* is the total contribution, the *dashed* (*red*) one the contribution of gauge fields, while in *solid* (*blue*) the contributions of the *top* and in *dot-dashed* (*green*) the light generations. Overall, the potential has a local minimum for small values of $\alpha \sim 0.2$

7 Second Model: A Minimal Composite Higgs, or Gauge-Higgs Unification in Warped Space

A warped extra dimension (or Randall-Sundrum space [40]) has been widely studied, because it can fairly easily generate hierarchies between mass scales. In models of Gauge-Higgs, it offers two main advantages: it automatically enhances both the Higgs and the top mass.

The difference between flat and warped space is the metric: the simple Minkowski metric in flat space is replaced by

$$ds^2 = e^{-2x_5/R} dx^\mu dx_\mu - dx_5^2\,, \quad x_5 \subset [0, L]\,. \tag{113}$$

This metric has an interesting property, conformal invariance, which is more evident if we rewrite it in terms of $z = Re^{x_5/R}$:

$$ds^2 = \left(\frac{R}{z}\right)^2 (dx^\mu dx_\mu - dz^2)\,, \quad z \subset [R, R' = Re^{L/R}]\,. \tag{114}$$

An increase in the value of z by a factor of ξ (i.e. $z \to \xi \cdot z$) can be compensated by an analogous rescaling of $x_\mu \to \xi \cdot x_\mu$ to leave ds^2 invariant. A simple and intuitive way to think about this symmetry is the following: changing z is equivalent to moving along the extra co-ordinate, which is then linked by the symmetry to a rescaling of the size (and therefore of the energy) of physical 4-dimensional systems. Different positions in the extra space correspond, therefore, to different energy scales. One can

choose the positions of the two end-points of the space so that $R^{-1} \sim M_{Pl}$ and $(R')^{-1} \sim 1\,\text{TeV}$: moving from the boundary at $z = R$ (Planck brane) to the $z = R'$ one (TeV brane) will rescale energy scales from the Planck scale down to the TeV. This symmetry is therefore considered to provide a solution of the hierarchy problem between the electro-weak and the Planck scales. Note that the length of the interval is $L = R \log R'/R$.

A gauge boson in the warped space will have an action [22]

$$\begin{aligned} \mathscr{S}_{gauge} &= -\tfrac{1}{4} \int d^4x\, dz\, \left(\tfrac{R}{z}\right)^5 F_{MN} F^{MN} \\ &= -\tfrac{1}{4} \int d^4x\, dz\, \left(\tfrac{R}{z}\right) (F_{\mu\nu} F^{\mu\nu} - 2F_{\mu z} F_z^{\mu}) \,. \end{aligned} \tag{115}$$

The factors of R/z come from the metric. As in the flat case, a gauge fixing term is added to remove A_μ-A_5 mixing:

$$\mathscr{S}_{GF} = -\frac{1}{2\xi} \int d^4x\, dz\, \left(\frac{R}{z}\right) (\partial^\mu A_\mu - \xi z \partial_z (A_5/z))^2 \,. \tag{116}$$

The equation of motion for the wave function of a vector is

$$z\partial_z \left(\frac{1}{z} \partial_z f_n\right) + m_n^2 f_n = 0 \,, \tag{117}$$

whose solutions can be expressed in terms of Bessel functions of the first and second kind:

$$f_n = z\,(A\, J_1(m_n z) + B\, Y_1(m_n z)) \,. \tag{118}$$

For the scalars, the equation of motion reads:

$$\partial_z \left(z \partial_z \left(\frac{A_5}{z}\right)\right) + \frac{m_n^2}{\xi} A_5 = 0 \,. \tag{119}$$

As before, massive mode are Goldstone bosons eaten by the massive vectors, while for the zero mode

$$A_5 \sim z \,. \tag{120}$$

7.1 Custodial Symmetry?

We may want to try constructing a SU(3) model: however this is not acceptable in warped space. The difference with respect to the flat case is that the BEH vacuum

depends linearly on the extra co-ordinate, thus mixing between various KK modes is possibly generated by it. In the flat case:

$$\langle A_5 \rangle W_n^+ W_m^0 \sim \langle A_5 \rangle \int f_n(x_5) f_m(x_5) = 0 \tag{121}$$

because the vacuum is a constant and two wave functions are orthogonal. In the warped case, the vacuum is not constant as $\langle A_5 \rangle \sim z$, therefore tree level corrections to the electro-weak precision measurements are usually generated, in particular to the ρ parameter. As we learned at the beginning of the lectures, the ρ parameters is protected against large corrections by the *custodial symmetry*: what we need here is therefore a way to implement the custodial symmetry in Gauge-Higgs models. The simplest way is to use SO(5) instead of SU(3) as the starting gauge group [3], because it contains SO(4) as a subgroup, and the coset space (which will embed the BEH field) is a bi-doublet of SO(4) – thus it has the correct custodial symmetry properties as the BEH field in the Standard Model. The extended gauge structure is then broken on the boundaries of the extra dimensions in the following way:

- On the TeV brane, we can break $SO(5) \to SO(4)$: the $SO(4) \sim SU(2) \times SU(2)$ contains the desired custodial symmetry (the breaking of this symmetry will be achieved via the BEH vacuum, like in the SM).
- The generators of SO(5) that do not belong to the unbroken subgroup SO(4) form a 4 of SO(4), like the BEH field in the SM!
- On the Planck brane, we break $SO(5) \to SU(2) \times U(1)$, so that only the SM gauge invariance is preserved at the level of zero modes. As it is a subgroup of the unbroken SO(4), only the $SU(2) \times U(1)$ gauge bosons have zero modes, as desired.
- For the scalars, only the BEH has a zero mode.

This structure of the symmetry breaking is enough to ensure that the values of the W and Z mass respect the SM relations at tree level (thus $\rho = 1$ at tree level). This symmetry structure can also be used to protect couplings of the Z boson to light fermions against large corrections [4].

7.2 AdS/CFT

The presence of a conformal symmetry in the metric suggests a correspondence between models in warped space (anti de Sitter) and strongly interacting conformal theories in 4 dimensions. The correspondence goes as follows:

- Fields and symmetries on the Planck brane correspond to the elementary sector of the theory (like the photon in QED+QCD);

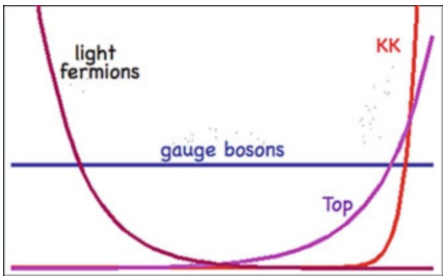

Fig. 4 Wave functions in the warped model for gauge bosons, light fermions, tops and KK modes, showing their localisation: the Planck brane is on the right, the TeV brane on the left. The Higgs is also moderately localised towards the TeV brane

– Fields in the bulk correspond to operators (bound states) of the conformal sector, the TeV brane breaks the conformal invariance and generates a mass gap (tower of meson resonances, where the mass gap corresponds to confinement);
– Symmetries in the bulk and on the TeV brane correspond to global symmetries of the strong sector (so, our strong sector is invariant under SO(5) which is spontaneously broken to SO(4)).

This model can therefore be seen as the SM (the Planck brane is invariant under the SM gauge group) coupled to a conformal sector which is invariant under a global SO(4) (that generates the custodial symmetry!). The BEH field is localised towards the TeV brane (as its wave function grown with z), therefore it is part of the strong sector and its vacuum solution is generated dynamically by the strongly interacting sector. Note that this mechanism is very similar to Technicolour, with the addition of the conformal symmetry. The correspondence has the status of a conjecture, without a formal mathematical proof: nevertheless, by studying the properties expected for a conformal strongly interacting sector and the properties of the fields in the warped space, one can define an accurate dictionary that relates the properties of the two sectors, both in the bosonic [17] and the fermionic sectors [13, 16].

The properties of all the fields depend on their localisation in the extra space: the cartoon in Fig. 4 shows the typical scenario. Gauge bosons have a flat profile (due to gauge invariance), while the Higgs is moderately localised toward the TeV brane thus revealing its composite origin. Light fermions, like leptons, light quarks and the right-handed bottom, are localised toward the Planck brane: they correspond to mostly elementary fields, and the localisation suppresses their overlap with the Higgs thus generating the hierarchies in the fermion mass spectrum. The top is localised toward the TeV brane, thus it is a mostly composite state: its localisation enhances the overlap with the Higgs, thus it makes possible to achieve masses larger than m_W, even though the coupling is of the order of the gauge couplings. The Higgs, being localised toward the TeV brane, is also a composite state! All the massive resonances (KK) are also strongly localised to the TeV brane, thus showing their composite nature.

7.3 Higgs Potential and Mass

The calculation of the Higgs potential proceeds as in the flat space, however the calculations are complicated by the presence of Bessel functions. Nevertheless, the result will look very similar to the Gauge-Higgs models in flat space, detailed in the previous section: the typical Higgs mass will range between $60 < m_H < 140$ GeV, thus the measured value can be easily obtained. More details on the impact of the Higgs mass measurements on this class of models can be found in [37].

8 Final Words

Many decades have passed between the first formulation of the Standard Model as a candidate to describe phenomena involving elementary particles, and now the last predicted particle has been discovered, the Higgs boson. This is, however, an open door rather than a lowering curtain for the field, because many questions are still waiting for answers: above all, the discovery of a scalar field is a materialisation of the naturalness problem. Despite the negative results of the LHC in the quest for New Physics, the last word is far from being spoken: new physics may reveal itself in ways that are difficult to disentangle from the abundant background in hadron colliders, and much more work is required from both the theoretical and experimental sides to finally pull the correct string and reveal the New Standard Model.

References

1. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012)
2. P.A.R. Ade et al., Planck Collaboration, arXiv:1303.5076 [astro-ph.CO].
3. K. Agashe, R. Contino, A. Pomarol, Nucl. Phys. B **719**, 165 (2005)
4. K. Agashe, R. Contino, L. Da Rold, A. Pomarol, Phys. Lett. B **641**, 62 (2006)
5. ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD and LEP Electroweak Working Group and Tevatron Electroweak Working Group and SLD Electroweak and Heavy Flavour Groups Collaborations, arXiv:1012.2367 [hep-ex]
6. I. Antoniadis, Phys. Lett. B **246**, 377 (1990)
7. I. Antoniadis, K. Benakli, Phys. Lett. B **326**, 69 (1994)
8. I. Antoniadis, K. Benakli, M. Quiros, New J. Phys. **3**, 20 (2001)
9. N. Arkani-Hamed, S. Dimopoulos, G.R. Dvali, Phys. Lett. B **429**, 263 (1998)
10. R. Barbieri, A. Strumia, hep-ph/0007265
11. G. Bertone, *Particle Dark Matter: Observations, Models and Searches* (Cambridge University Press, Cambridge, 2010)
12. G. Cacciapaglia, C. Csaki, S.C. Park, JHEP **0603**, 099 (2006)
13. G. Cacciapaglia, G. Marandella, J. Terning, JHEP **0906**, 027 (2009)
14. S. Chatrchyan et al., CMS Collaboration, JHEP **1306**, 081 (2013). arXiv:1303.4571 [hep-ex]

15. R. Contino, Physics of the large and the small, in *Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 09)*, Boulder, 1–26 June 2009 (Word Scientific, Singapore/Hackensack, 2009). arXiv:1005.4269 [hep-ph]
16. R. Contino, A. Pomarol, JHEP **0411**, 058 (2004)
17. R. Contino, Y. Nomura, A. Pomarol, Nucl. Phys. B **671**, 148 (2003)
18. C. Csaki, C. Grojean, H. Murayama, Phys. Rev. D **67**, 085012 (2003)
19. C. Csaki, C. Grojean, H. Murayama, L. Pilo, J. Terning, Phys. Rev. D **69**, 055006 (2004)
20. C. Csaki, C. Grojean, L. Pilo, J. Terning, Phys. Rev. Lett. **92**, 101802 (2004)
21. C. Csaki, C. Grojean, J. Hubisz, Y. Shirman, J. Terning, Phys. Rev. D **70**, 015012 (2004)
22. C. Csaki, J. Hubisz, P. Meade, Physics in D >= 4, in *Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 04)*, Boulder, 6 June–2 July 2009. (Word Scientific, Singapore/Hackensack, 2006). hep-ph/0510275
23. J.R. Ellis, M.K. Gaillard, D.V. Nanopoulos, Nucl. Phys. B **106**, 292 (1976)
24. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
25. G.F. Giudice, A. Masiero, Phys. Lett. B **206**, 480 (1988)
26. S.L. Glashow, Nucl. Phys. **22**, 579 (1961)
27. Y. Grossman, M. Neubert, Phys. Lett. B **474**, 361 (2000)
28. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
29. P.W. Higgs, Phys. Lett. **12**, 132 (1964)
30. P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
31. G. Hinshaw et al., WMAP Collaboration, Astrophys. J. Suppl. **208**, 19 (2013). arXiv:1212.5226 [astro-ph.CO]
32. Y. Hosotani, M. Mabe, Phys. Lett. B **615**, 257 (2005)
33. M. Kobayashi, T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973)
34. K.D. Lane, in Proceedings, The building blocks of creation, Boulder (1993), pp. 381–408, and Boston U. - BU-HEP-94-02 (94,rec.Jan.) 32p. [hep-ph/9401324]
35. S.P. Martin, in *Perspectives on supersymmetry II*, ed. by G.L. Kane (World Scientific, Singapore, 2010), pp. 1–153. [hep-ph/9709356]
36. G. Panico, M. Serone, A. Wulzer, Nucl. Phys. B **739**, 186 (2006)
37. G. Panico, M. Redi, A. Tesi, A. Wulzer, JHEP **1303**, 051 (2013)
38. B. Pontecorvo, Sov. Phys. JETP **6**, 429 (1957); Zh. Eksp. Teor. Fiz. **33**, 549 (1957)
39. S. Raby, in *Proceedings to the 2nd World Summit: Physics Beyond the Standard Model*, Galapagos Islands, 22–25 June 2006. hep-ph/0608183
40. L. Randall, R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999)
41. A.D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967); JETP Lett. **5**, 24 (1967); Sov. Phys. Usp. **34**, 392 (1991); Usp. Fiz. Nauk **161**, 61 (1991)
42. M. Schmaltz, D. Tucker-Smith, Ann. Rev. Nucl. Part. Sci. **55**, 229 (2005)
43. C.A. Scrucca, M. Serone, L. Silvestrini, Nucl. Phys. B **669**, 128 (2003)
44. M.F. Sohnius, Phys. Rept. **128**, 39 (1985)
45. J.W.F. Valle, J. Phys. Conf. Ser. **53**, 473 (2006)
46. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967)

Part II
Experimental Results from the Large Hadron Collider

Probing the Standard Model at Hadron Colliders

Peter Mättig

Abstract Although the Standard Model has been confirmed to stunning precision, at each new collider it is further scrutinized to ever-increasing precision. Measurements of Standard Model processes are an essential part of the physics program at the Tevatron and LHC proton colliders. Apart from challenging the theory in new energy regimes and phase spaces, they provide the means to search for extensions of the Standard Model and give confidence in the tools used for new physics searches. Indeed, deviations from Standard Model predictions may be the first indications for physics beyond the Standard Model. In these lectures, the experimental methods used at the LHC and Tevatron will be discussed and results presented on QCD and electroweak processes, as well as on the study of properties of the top quark.

1 Why Standard Model-Test at the LHC

With the discovery of the Higgs boson at a mass of around 125 GeV [1, 2], all parts of the Standard Model have been observed. Moreover, most of its elements have been scrutinised with very high accuracy, making the Standard Model probably the most precise and comprehensive theory in the history of physics. Given this enormous progress of the last decades, what is there to add by the current highest-energy hadron colliders? Physics at the Tevatron and LHC proceed in at least three main directions:

1. Exploring processes not determined from first principles, a notable example being soft hadron interactions;
2. Improving measurements on basic elements of the Standard Model, a prime example being the top quark;
3. Probing the Standard Model at unexplored energy regimes, like jet cross sections at the deep TeV scale.

P. Mättig (✉)
Bergische Universität Wuppertal, Wuppertal, Germany
e-mail: peter.mattig@cern.ch

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_4

Today's High Energy Physics is characterised by the quest to find 'New Physics' in response to aesthetic deficiencies of the Standard Model and hints from astrophysical observations. Measurements of Standard Model processes are an integral part of this program. They are backgrounds to New Physics signatures and provide tools to study those. Moreover, the first indications of a breakthrough towards New Physics may be seen in deviations from Standard Model expectations. Studying Standard Model processes, especially in the highest energy regimes, is thus indistinguishable from searching for Beyond the Standard Model effects.

2 Experimentation at the LHC

Until recently, the Tevatron at Fermilab (Batavia/USA) was the frontier collider with $p\bar{p}$ energies of 1.96 TeV. Its two experiments CDF and D0 each collected around 10 fb^{-1} over the last 10 years. Since 2010, the LHC has been producing pp collisions at 7 and 8 TeV. Due to outstanding performance the LHC has delivered close to 30 fb^{-1} with a peak luminosity of close to 8×10^{33} $\text{m}^{-2}\text{sec}^{-1}$. Higher energies, higher luminosities and state-of-the-art detectors give the LHC the lead in particle physics. Yet the Tevatron remains an important complement including measurements of the highest precision and somewhat different production processes for energies of 300–1,500 GeV.

Four large detectors are built around the points where pp collide inside the LHC ring. Whereas ALICE's specific aim is to understand heavy ion collisons and LHCb is specialising in bottom and charm production, the largest experiments ATLAS [3] and CMS [4] are to explore the highest energies and to study Standard Model processes as well as search for phenomena beyond. The results of these two experiments will dominate these lectures.

Both detectors are huge and structured in layers for measuring charged particles (tracking detectors), electromagnetically interacting particles (electromagnetic calorimeter) and hadronically interacting particles (hadron calorimeter). At the very outside of both experiments tracking chambers are added to identify and measure muons. Both experiments cover almost the total solid angle around the pp interaction point leaving just 14–15 mrad uncovered to make room for the beam pipe. ATLAS is the larger detector with a length of 45 m and a diameter of 25 m. It weighs some 7,000 tons. CMS is the heavier experiment of 12,500 tons being 'only' 21.6 m long and 14.6 m in diameter. Each of these experiments has some 3,000 authors on their publications. In spite of the size of these detectors, they are filled to the last cm^3 with highly sophisticated devices to provide extremely precise measurements. Whereas the global structures of the experiments are the same, there are important differences in detail.

2.1 Tracking at ATLAS and CMS

The tracking detectors of the two experiments extend over radii of about 50–1,000 mm from the beam line. CMS uses only silicon detectors. In its central part, there are 3 layers of pixel detectors close to the collision point, complemented at larger radii by 7 layers of strip detectors leading to a total of 17 possible space points for each particle traversing the central part of the detector. These detectors are immersed in a magnetic field of 4 T. The barrel part of the ATLAS tracking detector also has 3 (soon 4) pixel layers, but only 4 strip layers, while a transition radiation tracker covers radii larger than 550 mm and provides up to 35 additional measurements (with an axial position resolution of 130 μm) and electron identification capability. The ATLAS tracker is inside a magnetic solenoid of 2 T. The forward regions of the two detectors are instrumented similarly to the barrel part.

The typical measurement accuracy for an individual point is 10–20 (20) μm for pixels (strips) in the plane transverse to the beam direction and some 50–100 (600) μm in the direction parallel to the beam such that the detectors provide excellent measurements of the momenta of charged particles. With somewhat simplified assumptions the resolution of tracking detectors in the plane transverse to the magnetic field is given by [5]

$$\frac{1}{p_T}\frac{\sigma(1/p_T)}{1/p_T} = \frac{1}{0.3\cdot B}\frac{\sigma_{point}}{L^2}\sqrt{\frac{720}{N_{point}+4}} \tag{1}$$

Here N_{point} equally spaced measurements are assumed, each with a precision of δ_{point}. The magnetic field in the transverse $(x,\ y)$ plane is B and the projected length in this plane is L. Although these conditions are not strictly true for the LHC experiments, key features become apparent. The resolution improves quadratically with L, linearly with B and δ_{point}, and only marginally with the number of measurements. Taking into account material effects, multiple scattering degrades the resolution to

$$\frac{\sigma(p_T)}{p_T} = a \oplus b\cdot p_T \tag{2}$$

where b is given by the right side of Eq. 1 and a is a constant term due to multiple scattering.

Whereas radii and point resolutions are fairly equivalent for the two experiments, the higher magnetic field B of CMS translates into a better momentum resolution. For muons the p_T resolution of CMS [6] is shown in Fig. 1 (left). The cross over between the multiple scattering term and the ideal resolution of Eq. 1 is at about 45 (90) GeV for ATLAS (CMS).

In addition to the momentum measurement, a highly important function of the tracking detectors is the identification of vertices. Both the position of the primary vertex, especially in the environment of many pile-up events (see Sect. 3.3), and

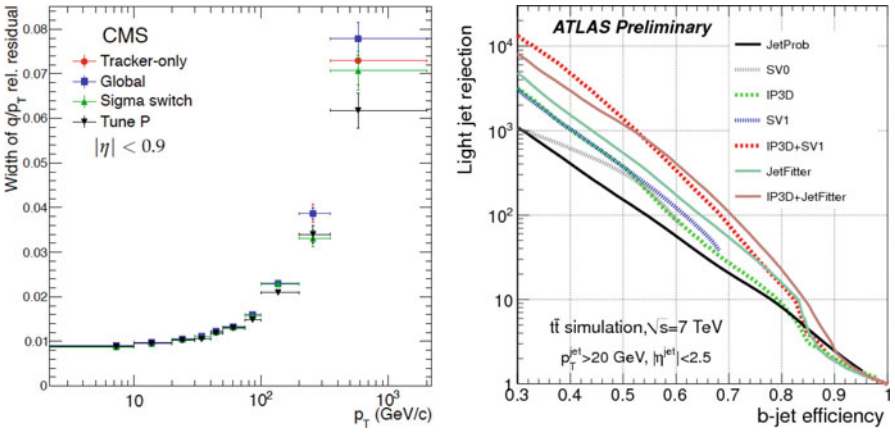

Fig. 1 Resolution of the transverse momentum of muons in the CMS detector [6] (*left*) and the bottom-tagging efficiency versus light quark rejection in ATLAS [7] (*right*)

the identification of a secondary vertex from a decay of a bottom hadron provide essential information for physics. Their precision is closely related to the one of the impact parameter d_0, i.e. the distance of closest approach to the primary interaction point. Assuming just two measurements at distances r_1, r_2 with point precisions $\sigma_{1,2}$ from the primary vertex, the precision is given by

$$\sigma^2(d_0) \;=\; \frac{\sigma_2^2 \cdot r_1^2 + \sigma_1^2 \cdot r_2^2}{(r_2 - r_1)^2} \tag{3}$$

This rewards a long lever arm and especially a very good point resolution close to the interaction point. Again, this simple relation is modified by multiple scattering, e.g. in the beam pipe, degrading the resolution to

$$\sigma^2(d_0) \;=\; \frac{c^2}{p_T^2} \;+\; d^2 \tag{4}$$

where d is given by Eq. 3 generalised to more measurements. The first term dominates the resolution for low energy particles, with $c \;\sim\; \sqrt{x/X_0}\;\; 1/r_{scat}$ depending on the r_{scat}, the distance of the scattering centre from the primary vertex and degrading with the amount of material in terms of the fraction of radiation length x/X_0. It is therefore important not only to minimise the material, but also to have a precise measurement as close as possible to the interaction point. The current ATLAS and CMS detectors provide efficient bottom tagging while strongly rejecting light flavours as shown in Fig. 1 (right) [7].

2.2 Calorimeters at ATLAS and CMS

Whereas tracking detectors are sensitive only to electrically charged particles, calorimeters are used to determine the energies also of neutral particles (except neutrinos). They measure the multiplicity of secondary particles in a 'shower': a chain of interactions induced by incoming particles. This number is essentially proportional to the energy of the primary particles. The properties of these showers are different for electromagnetically and for hadronically interacting particles: the former being rather narrow and short; the latter fairly broad, extended and having much larger fluctuations. To account for these differences, the calorimeters are separated into fine granularity compartments to identify electromagnetically interacting photons and electrons, and coarser ones to measure hadrons.

In the case of the electromagnetic calorimeter, ATLAS and CMS follow different concepts, see Fig. 2. ATLAS uses a sampling calorimeter, i.e. alternating layers of passive and active elements, with longitudinal and fine lateral granularity of up to $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. This allows the shower propagation to be followed. CMS uses homogeneous $PbWO_4$ crystals. Their very small radiation length is used for an even smaller lateral granularity of 0.017×0.017, at the price of having no longitudinal segmentation. Without material in front of the calorimeters, the resolutions for ATLAS and CMS are:

$$\frac{\sigma(E)}{E} = \frac{0.1}{\sqrt{E}} \oplus 0.002 \quad \text{(ATLAS)} \quad ; \quad \frac{\sigma(E)}{E} = \frac{0.028}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.003 \quad \text{(CMS)}. \tag{5}$$

As an example, these correspond to mass resolutions on a 126 GeV Higgs boson in its decay into two photons in the central region of 1.37 and 1.17 GeV respectively.

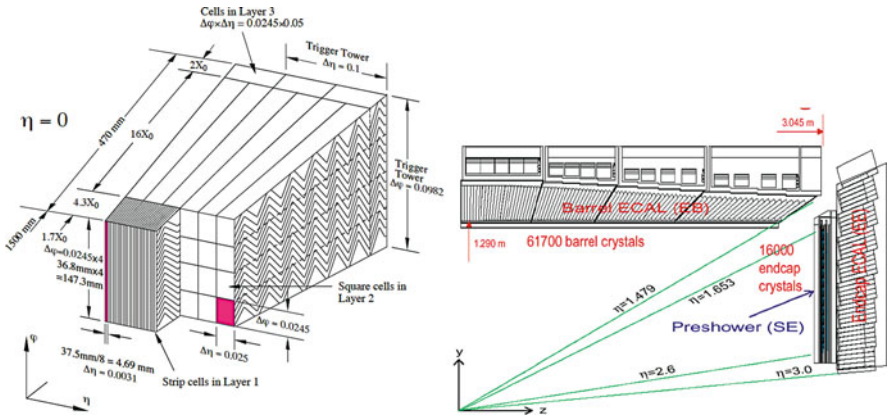

Fig. 2 Structure of the electromagnetic calorimeters of ATLAS (*left*) and CMS (*right*) (Taken from [4] and [3])

As hadronic calorimeters, both experiments use sampling calorimeters with either iron (ATLAS) or brass (CMS) as passive materials and scintillators as active materials in the central region. Their respective granularities are 0.1×0.1 and 0.087×0.087 and the resolutions in the central region are:

$$\frac{\sigma(E)}{E} = \frac{0.5}{\sqrt{E}} \oplus 0.055 \quad \text{(ATLAS)} \quad ; \quad \frac{\sigma(E)}{E} = \frac{1}{\sqrt{E}} \oplus 0.05 \quad \text{(CMS)} \tag{6}$$

In the forward region, ATLAS uses copper, and CMS tungsten and liquid argon.

2.3 Muon Detection at ATLAS and CMS

At the very outside of the detectors, tracking chambers are installed to catch muons, the only charged particles traversing the calorimeters. Two kinds of chambers are installed: one with a fast response for triggering, and the second to provide accurate offline reconstruction. To make use of their intrinsic high resolution, the chambers are supplemented by sophisticated alignment systems to monitor the chamber positions to a few tens of microns.

The chambers are embedded in a magnetic field provided either by special huge toroidal magnets as in ATLAS, or the magnetic return yoke for CMS with standalone resolutions of

$$\frac{\sigma p_T}{p_T} = 0.03 - 0.04 \quad \text{(ATLAS)} \quad ; \quad 0.1 \quad \text{(CMS)} \tag{7}$$

The muon p_T resolutions are significantly improved at least for momenta up to several tens of GeV, if the measurements of the muon system and the tracking detector are combined.

2.4 Data Selection and Reconstruction at ATLAS and CMS

The huge amount of data for all interactions and potential read-out inefficiencies require data selection to be made within a few micro- or milliseconds. Trigger systems are based on either special electronic cards or computer farms. ATLAS employs a three level system based on a 'region of interest' given by the Level 1 trigger. CMS merges all regions at an early stage and uses only a Level 1 decison together with the event builder. For both experiments, the Level 1 decision is based on muon chambers and calorimeter information. Some 300–400 Hz out of the 40 MHz of collision rate are written to disk.

Even with this reduction by a factor 100,000, the amount of data that is stored is huge. Together with the simulation required, some 25 Pb of data per year and experiment are stored. Furthermore, sophisticated algorithms are invoked to

translate the electronic signals into physics objects. This led to the development of Grid Computing, currently comprising around 100,000 CPU cores and 180 PB of disk space. These resources are distributed over some 100 institutes all over the world.

The performance goals of both the ATLAS and CMS detectors have been fully met and have sometimes even superseded expectations. This is the basis for the physics output of both experiments.

3 How Protons Interact

Probing high masses and correspondingly tiny regions of space requires a large energy density. The protons of the LHC are just vehicles for its subcomponents to have pointlike collisions that can be of either quark–quark, quark–gluon, or gluon–gluon type, and will be denoted as parton scattering.

3.1 Contributions to the Cross Section

The production of a system X by two partons q, q' of flavours f, f' is given by

$$\sigma(pp \to YX) = \int_0^1 dx \int_0^1 dx' \sum_{f,f'} F_f F_{f'} \sigma(p_f(xP) + p_{f'}(x'P) \to X) \cdot \sigma(p_r p'_r \to Y) \tag{8}$$

The state X will then decay into Standard Model particles, mostly quarks and gluons. The components of this formula are:

- $x = E/P$ (for $P \to \infty$), the energy fraction of a parton relative to the proton energy P.
- $F_f(x, Q^2)$ are the parton distribution functions (pdfs) describing the probabilities for a parton of flavour f to have an energy fraction x. These F_f depend on Q^2, an energy scale that characterises the hard scatter, e.g. $Q^2 = x \cdot x' P^2$ or the p_T of the outgoing partons.
- $\sigma(p_f(xP) + p_{f'}(x'P) \to X)$ is the cross section of the scattering of the two partons to produce a system X.
- Y characterises properties of the 'underlying event', i.e. the interactions of the residual parts of the proton p_r, which are not directly involved in the hard scatter.

Here it is assumed that the production of X and Y factorize. The energy fractions x_i should be high enough to produce the system X of mass M. Since

$$M = \sqrt{x_1 x_2} \cdot E_{pp} \tag{9}$$

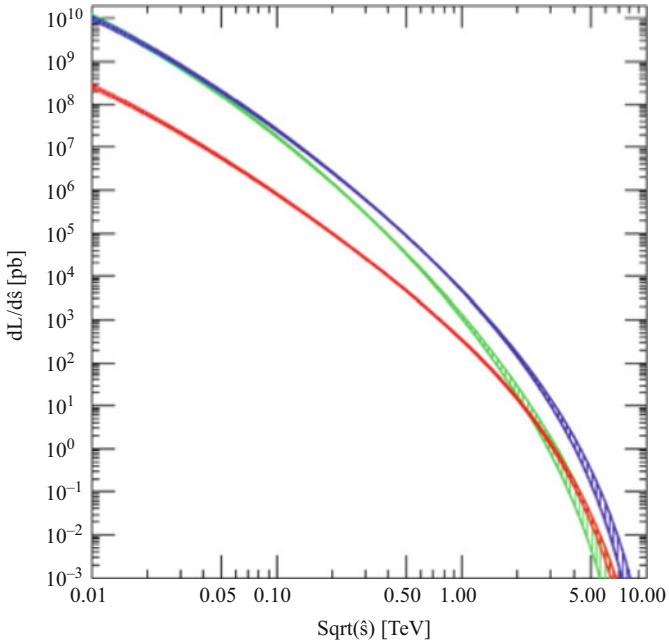

Fig. 3 Parton – parton luminosities to produce a system with $\sqrt{\hat{s}} = M$ at E_{pp} of 14 TeV [8]

high masses require high x_i. Here E_{pp} is the total pp – invariant mass, 1.96 at the Tevatron, currently 8 TeV at the LHC. Whereas at low M/E_{pp} gluon scattering dominates, very high masses are mostly due to quark scattering. This means that for masses of 0.3–1.5 TeV parton interactions at the Tevatron and the LHC are mostly of different type and their measurements complementary. The parton – parton luminosities for gg (green) and qq (red) are depicted in Fig. 3 for $E_{pp} =$ 14 TeV [8].

3.2 Observables at the LHC

In addition to their Q^2, collisions are characterized by their angular distributions. In proton collisions these are expressed in terms of rapidity and pseudorapidity. The rapidity is

$$y = \frac{1}{2}\ln\left(\frac{E + p_{||}}{E - p_{||}}\right) = \frac{1}{2}\ln\left(\frac{E + p_{||}}{\sqrt{M^2 + p_T^2}}\right) \tag{10}$$

where $p_{||}$ is the momentum component parallel to a boost direction, in general the beam direction, and p_T is the transverse component. The Lorentz transformation of the rapidity is given by

$$y \rightarrow y' = y + \frac{1}{2} \ln \left(\frac{1+\beta}{1-\beta} \right) \tag{11}$$

For a decay $X \rightarrow a, b$, it follows that $y_a - y_b$ is invariant under a boost β along the beam direction, the relevant direction of parton collisions. Therefore also ΔR_y, the measure of the distance between a and b, is Lorentz invariant:

$$\Delta R_y = \sqrt{(y_a - y_b)^2 + (\phi_1 - \phi_2)^2} \tag{12}$$

Frequently, the rapidity is replaced by the pseudorapidity η, with $y \rightarrow \eta$ for $m \rightarrow 0$

$$\eta = \frac{1}{2} \ln(\tan \theta/2), \tag{13}$$

which is just given by the scattering angle. Even if $\eta_1 - \eta_2$ is not Lorentz – invariant, the distance measure ΔR is frequently preferred over ΔR_y of Eq. 12:

$$\Delta R = \sqrt{(\eta_a - \eta_b)^2 + (\phi_a - \phi_b)^2} \, . \tag{14}$$

3.3 The Global Structure of Proton Collisions

Quarks and gluons are produced in most hard processes at the LHC. Their experimental signatures are jets: bundles of hadrons representing the original partons. Apart from those from the hard process, additional partons are radiated from initial and final state partons. Furthermore, there is a spray of particles, the 'underlying event', due to the colour flow between the proton remnants. The experimental challenge is thus to relate the observed jets to primary partons, to extract those due to the process of interest, and to reduce the impact of the underlying event on the jet measurements.

Precise calculations for the hard processes from first principles require perturbative expansions. Even at $Q^2 = M_Z^2$ the strong coupling is [9] is fairly large:

$$\alpha_s(M_Z^2) = 0.1184 \pm 0.0007, \tag{15}$$

such that good precisions require next-to-leading (NLO) or NNLO calculations. The coupling α_s has a strong Q^2 dependence, and 'soft' QCD interactions at low Q^2 are in a regime where α_s is large and perturbative methods no longer apply. Therefore, instead of calculating parton or particle distributions from first principles, assumptions and short-cuts have to be applied.

Table 1 Three classes of measurement at proton colliders

Underlying event	Parton distribution function	$\sigma(p_f(xP) + p_{f'}(x'P) \to X)$
Measure	From previous measurement	From theoretical calculation
From models	**Measure**	From theoretical calculation
From models	From previous measurement	**Measure**

Additional complications occur in the measurement due to many simultaneous proton–proton interactions named 'pile-up' interactions. Their number is proportional to the luminosity per bunch crossing and increases slowly with the pp energy. For the 2012 LHC running, typically 30 pile-up events are produced. These simultaneous interactions affect the event reconstruction.

Following Eq. 8, physics analyses can be separated into three complementary classes as shown in Table 1. These ingredients are only mildly related in the analyses, although in measuring e.g. the hard process, the limited knowledge of pdfs and the underlying events introduce systematic uncertainties. A full understanding requires some iterative procedures. These lectures will start discussing soft interactions before turning to the centrepiece of LHC physics: the hard interactions. The pdfs will only be mentioned in passing, since they have been addressed in a special lecture at this School.

4 Soft QCD Interactions

The physics of soft QCD interactions is to a large extent characterized by the fairly uniform production of a large number of particles. Studies of them provide insight into processes that cannot be calculated from first principles but are important to extract high Q^2 processes.[1] Soft QCD leads to the vast majority of events at proton colliders and contributes to the particle and energy distribution in events with hard interactions.

4.1 Models of Soft QCD

Because of the low Q^2 of these processes, α_s becomes large and a perturbative treatment impossible. Instead, QCD motivated concepts are introduced for both the hadron content and properties inside jets and the particle production outside jets. The former has been studied in detail in e^+e^- collisions [10] and are described in models like PYTHIA [11] and HERWIG [12]. Because of the remaining parts of

[1] They also inform the interaction between particle and astro-particle physics.

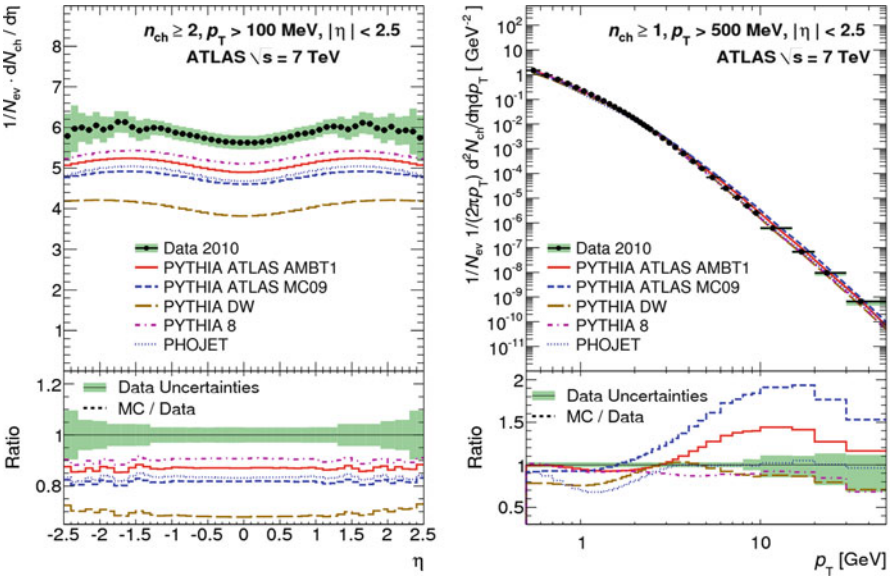

Fig. 4 Properties of minimum bias events (*left*) charged particle multiplicity as a function of pseudorapidity, (*right*) p_T spectrum [14]

the protons that do not directly participate in the hard interaction, for pp collisions additional features have been introduced. There are various approaches to model these as has been discussed at this School [13].

These models require parameters to describe the event properties. These have to be determined by data, however suffer from ambiguities which are taken into account by several 'tunes'. A lot of progress has been reached during the last decades, but it is unclear how precisely event properties can finally be described.

4.2 Minimum Bias Events

The overwhelming majority of pp interactions are soft, and the impact parameter of the protons small. They can be studied with a very loose trigger – essentially just indicating that indeed a proton–proton interaction has taken place – and are known as 'minimum bias events'. They have been studied in the first year of LHC data-taking and provide substantial insights into the models of soft QCD interactions [14, 15].

As can be seen from Fig. 4 (left), at a pp energy of 7 TeV, some 6 charged particles of $p_T > 100$ MeV are produced in the central region per $\Delta\eta = 1$. Figure 4 (right) shows that most of these particles have fairly low momenta. Tunes of

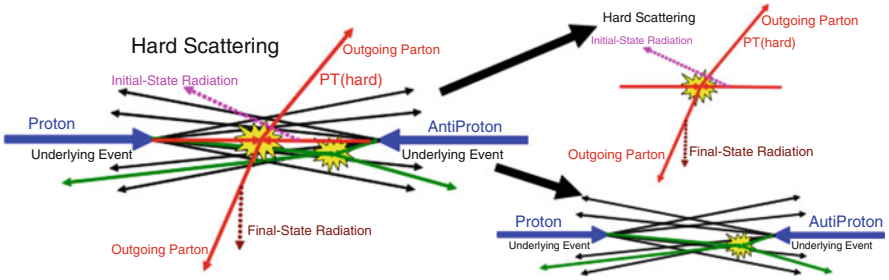

Fig. 5 Schematic of the underlying event [17]

the model before LHC data taking show an insufficient description of the data. By now, adjustments have been made to account for the LHC results.

These minimum bias events are identical to the pile-up events in hard collisions and therefore crucial to understanding how the reconstruction of the hard processes is affected: for example, the measurements show that for 2012 LHC luminosities some 1,000 charged particles from pile-up are in the tracker volume.

A special category of pp scattering is double hard parton scattering, where two instead of just one parton inside each colliding proton interact. Measurements of their rates have been performed at the Tevatron and the LHC [16].

4.3 Underlying Events

A schematic view of an underlying event is depicted in Fig. 5. Its properties are studied in events with two jets and in those with W and Z bosons. These studies are performed at low luminosities where pile-up events do not disturb the measurements.

As known from the clean environment of e^+e^- collisions, two jet events only have small contributions in the directions perpendicular to the jet axis. This is the motivation to study the properties of the underlying event in pp collisions [18] by dividing the plane perpendicular to the beam axis into four quadrants. One quadrant is centered along the axis of the leading (highest p_T) jet. The quadrants perpendicular to this axis are expected to be only minimally affected by the hard scattering and instead reflect the properties of the underlying event. The other method of using W and Z bosons is more simple. Since these vector bosons are colour neutral, all particles other than their leptonic decay products are due to the underlying event.

Figure 6 shows the track density in two-jet events as a function of the p_T of the leading jet, and the energy density in Drell-Yan production as a function of the p_T of the vector boson, which reflects the hardness of the process [19, 20]. The density of the track multiplicity (left) is fairly independent of the jet p_T, the energy

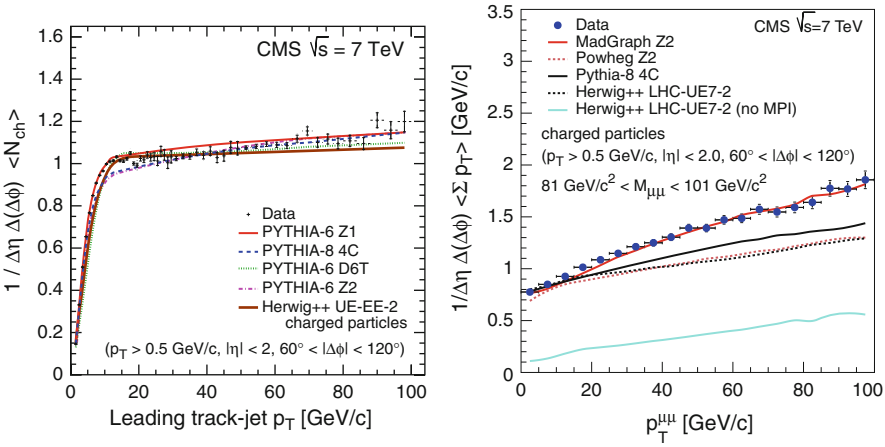

Fig. 6 Particle density (*left*) in the transverse quadrant of 2 jet events [19] and (*right*) in Drell Yan events [20]

density (right) shows a linear rise with the p_T. Whereas the detailed structure of an underlying event depends on the structure of the hard interactions, these measurements can gauge the modelling and allow the free parameters to be tuned. For example, the track density is rather well described by PYTHIA with various tunes; but on the other hand the size of the energy density and its dependence on $p_T^{\mu\mu}$ is unsatisfactory.

5 Hard QCD Interactions

Comparing theoretical predictions for hard interactions with experimental results requires as precise as possible knowledge of the parton distribution functions. Measurements from previous experiments, notably the high precision data from the electron – proton collider HERA [21] but also from several types of LHC measurements provide input to QCD motivated parton distribution for different flavours and Q^2. Details have been covered at this School.

5.1 Finding Jets

Hard QCD processes at the LHC lead to quarks and gluons jets. These jets are universal, i.e. they have the same properties in e^+e^-, ep or pp collisions, allowing results from one type of process to be applied to another.

The definition of a jet is not unique. Several procedures for reconstructing jets exist and all contain parameters that may be optimised according to the physics process in question. To allow for comparison with theory, and for well-understood experimental corrections to be made, a few requirements define a good jet finder:

- It is collinear safe, i.e. splitting a parton into two collinear ones does not change the jet properties;
- It is infrared safe, i.e. adding an (infinitely) low energetic parton does not change the jet properties;
- It facilitates experimental corrections for 'noise', e.g. contributions from other sources than the hard interaction as those due to pile-up events.

Whereas over decades the favourite jet finding algorithm at proton colliders has been the cone jet-finder, at the LHC in most cases the 'Anti-kt' finder [22] is used. This jet finder fulfils all of these requirements. It sequentially combines particles into proto-jets according to the similarity measure

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \quad \text{with} \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2. \quad (16)$$

Therefore, jet formation starts with the particles of $\min(1/p_{Ti}^2)$, i.e. with high p_T particles. This is followed by assigning particles which are spatially close, i.e. with a small ΔR. Those are mostly low p_T particles. The construction of the jet terminates if $\Delta R_{ij}^2 = R^2$, which has predefined values of typically 0.4 or 0.5. These jets are fairly circular in (η, ϕ). Next, another jet is constructed according to this algorithm until all particles not too close to the beam axis are assigned to jets.

The sum of the momenta of all particles inside a jet is taken as the momentum of the original parton. In addition, experimental effects have to be accounted for. Since jets are a mixture of electromagnetically and hadronically interacting particles, of neutral and electrically charged ones, each contributing different signals in the various compartments of the detector, the jet energy resolution relies on a clever combination of the available information. ATLAS and CMS approach this from different angles reflecting their different detector concepts.

ATLAS uses the calorimeter information alone and sums up the adjacent energy depositions in the electromagnetic and hadronic segments to form clusters. It makes use of the very good lateral and longitudinal segmentations allowing the identification of energy hot spots inside the hadronic shower due to its electromagnetic component. This is needed since the different kinds of particles have a somewhat different energy deposition (e/π ratio), even though the difference is relatively small for ATLAS.

CMS uses a 'particle flow' algorithm which combines their precise measurements of momenta in the tracking detector with the energy depositions observed in the calorimeter [23]. Since 50–60 % of the jet energy is due to charged particles, this procedure is particularly useful, if the tracking precision is superior to the calorimetric one. Tracking information is also the only way in which particles can be assigned reliably to a vertex. It may therefore help in compensating for distortions from the large number of pile-up events.

The parton flavour is more difficult to reconstruct with the notable exception of the bottom quark (cf. Sect. 2.1). Methods to identify others like charm quarks and to separate gluons and quarks are possible but suffer from either low efficiencies or purities and are rarely used. The precision of reconstructing energy and directions of the partons is worse than those of leptons.

5.2 Jet Cross Section

The immediate QCD test at hadron colliders is the measurement of the jet cross section. It is theoretically predicted to NLO and the LHC probes QCD in completely new energy regimes. The main experimental challenges are to correctly relate the electronic signals in the detector to the parton energy and to subtract contributions from the noise of pile-up events.

A small uncertainty in the energy scale is magnified in the uncertainty of the cross section. The latter is steeply falling with p_T like

$$\frac{d\sigma}{dp_T} \propto e^{-a\ p_T} \tag{17}$$

with $a \sim 0.04\,\text{GeV}^{-1}$. Thus a scale uncertainty of 2.5 % translates into an uncertainty in the cross section of 10 % at 100 GeV.

Several methods are used to calibrate the energy scale. A standard one uses events where a single jet recoils against a photon. Since the photon energy can be measured rather precisely, the jet p_T follows from the p_T balance. Depending on the energy the uncertainty varies from close to 3 % at $p_T = 25\,\text{GeV}$ to less than 1 % for energies of several hundred GeV [24]. There are several individual sources contributing to this total uncertainty, like the photon energy scale, QCD radiation effects, and the fraction of the parton energy that is sprayed outside the jet.

Another experimental task is to understand the energy resolution which makes a steeply falling function look flatter. This is because an event with a 'true' energy E_1 with a high cross section can migrate to an observed energy E_2 which has a smaller cross section. This loss at E_1 is not compensated by the reverse migration from the true E_2 to the measured one of E_1. The energy resolution can be determined from the p_T balance in two-jet events. Once known, it can be used to unfold the measured distribution.

Since contributions from the high number of pile-up events cannot easily be distinguished from those of the hard interactions, their energy depositions will be inadvertedly included in jet finding. This distorts measurements of the jet energy and direction along with other quantities like the missing energy of a hard scatter. For example, for anti-kT jets of $R = 0.5$, CMS finds an additional 0.5 GeV in a jet per pile up event [25]. An event-wise correction can be applied by the 'jet area

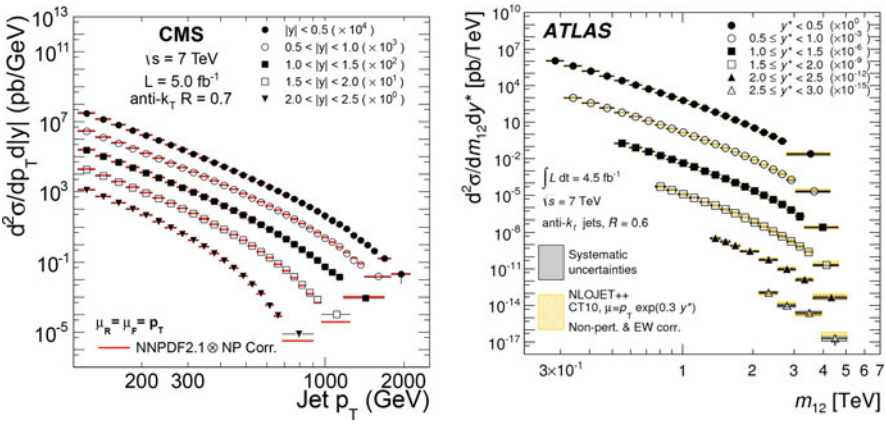

Fig. 7 Cross sections of jet production corrected to hadron level. *Left*: CMS measurement for various rapidity intervals as a function of p_T [26]. *Right*: the ATLAS measurement of the jet – jet mass for various rapidities [27]. In both cases the measured data are compared to a NLO calculation convolved with non – perturbative effects

method'. It is based on the homogeneity of the energy deposition of pile-up events. The energy density $\rho = E/\Delta R$ outside the hard components of the interaction is determined for each individual event which in turn is used to subtract an energy

$$E_{subtract} = A \cdot \rho \tag{18}$$

from the jet, where A is its area. Since a jet consists only of discrete depositions, the determination of A requires a special procedure. One way is to add 'infinitely many additional' particles of tiny energy homogeneously across to the $\eta - \phi$ plane and determine which ones are assigned to which jet.

The measured cross sections as a function of jet p_T [26] and the jet–jet mass [27] are displayed in Fig. 7. They probe QCD over nine orders of magnitude up to unprecedented mass scales of several TeV. The total uncertainties in the jet cross sections are of $\mathcal{O}$(10–20 %). They are largest for the very low and very high p_T and are dominated by the jet energy scale. QCD predictions describe the data amazingly well over this huge range. No apparent deviation is observed in these inclusive distributions.

These cross sections at the multi – TeV range can also be used to search for New Physics. For example, new jet–jet resonances from an excited quark decay $q^* \rightarrow qg$ would appear as an enhancement at the resonance mass. Since no significant signal is seen, a limit of $M_{q^*} > 3.6$ TeV [28] can be set.

Parton substructures, on the other hand, would lead to a steady rise or depletion of the cross section compared to the Standard Model prediction. As a convention such a deviation is parameterized by contact interactions [29], which could have different helicity structures η_{ij}, with L for left, R for right-handed coupling, e.g.

$$\mathscr{L}_{CI} \;=\; \frac{g^2}{\Lambda^2}\eta_{LL}(\bar{\psi}_L\gamma^\mu\psi_L)(\bar{\psi}_L\gamma_\mu\psi_L) \;+\; (RR, LR) \tag{19}$$

This formalism is shaped after Fermi's four – fermion theory of weak interactions. The effective compositeness scale is denoted by Λ and the strength by the coupling $g^2 \;=\; 4\pi$. The contact interaction interferes with the Standard Model contribution leading to

$$\sigma_{ff} \;=\mid \mathscr{M}_{SM} \mid^2 \;+ 2\frac{1}{\Lambda^2}\Re(\mathscr{M}_{SM}\cdot\mathscr{M}_{CI}) \;+\; \frac{1}{\Lambda^4} \mid \mathscr{M}_{CI} \mid^2 \tag{20}$$

where $\mathscr{M}$ denotes the relevant matrix element. As in the Fermi theory, $\mathscr{M}_{CI}$ grows with $\sqrt{s}$ such that for $\Lambda \gg s$, the last term is negligible and the interference term dominates possible deviations from the Standard Model, which grow like

$$\sigma_{CI}/\sigma_{SM} \;\sim\; \frac{M_{jj}^2}{\Lambda^2} \quad \text{for} \quad M_{jj}^2 \ll \Lambda^2 \tag{21}$$

Such a deviation could become visible even for masses much lower than the compositeness scale.

The sensitivity to new effects is increased by measuring the angular distribution of the outgoing partons. The difference in the rapidities of two decay products is related to the decay angle. Choosing

$$\chi \;=\; e^{|y_1 - y_2|} \tag{22}$$

a flat distribution is expected from the Standard Model since a high mass jet–jet system has a large component due to the t-channel contribution, leading to jets that are predominantly forward – backward going. New particles, on the other hand, decay more isotropically, i.e. tend to have a small Δy and $\chi \;\sim 1$.

The measurement is shown in Fig. 8. Contact interactions with a scale of up to 7.8 TeV, as well as colour octet quarks, quantum black holes etc. can be excluded [28].

5.3 Structure of Jets

5.3.1 Gluon Radiation

Beyond the inclusive jet cross section, QCD dynamics becomes visible through gluon emission. which leads to an increased number and acollinear jets. These effects have been measured at the Tevatron and the LHC and compared to QCD calculations. As an example, the angle between the highest p_T jets in the plane orthogonal to the beam axis is shown in Fig. 9. This acollinearity reflects the

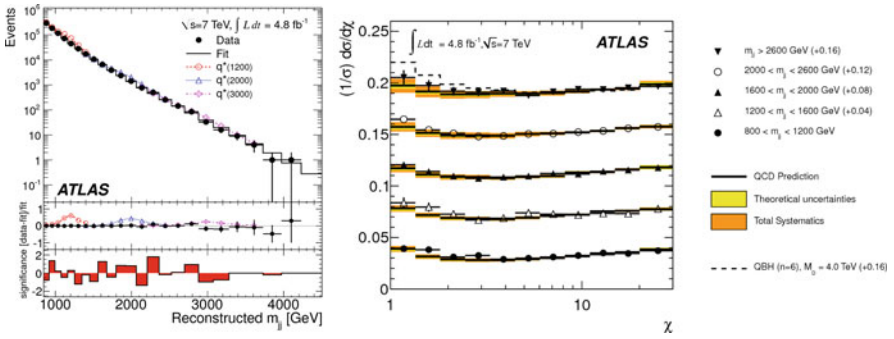

Fig. 8 Searches for BSM effects in jet – related cross sections [28]. *Left*: ATLAS measurement of the observable χ (see text) for various mass intervals. Also shown are the expectations from QCD and for a Quantum Black Hole. *Right*: the ATLAS measurement of the jet–jet mass compared to the expectation for an excited quark decaying into qg

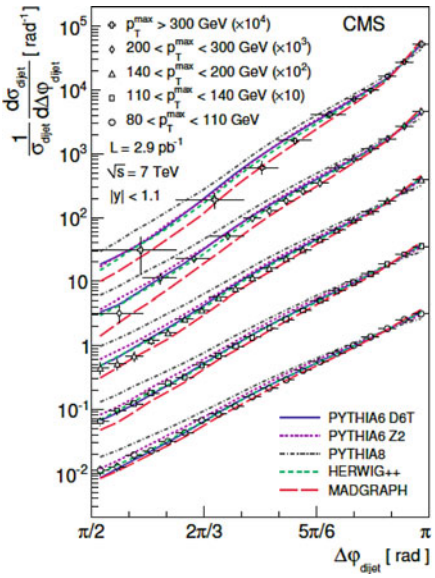

Fig. 9 Measurement of the angle $\Delta\phi_{dijet}$ between the highest energy jets in the transverse plane to the beam direction [30]

hardness of gluon emission [30]. It is shown for various maximum jet p_T and compared to some QCD calculations. In general the trend is reproduced by all models; however, at around $\pi/2$ some PYTHIA tunes overestimate the yield by several 10 %.

Such results can also be used to determine α_s. For example, the D0 collaboration has counted the average number of neighbouring jets within a predefined ΔR as a function of Q^2 and obtains [31]

$$\alpha_s(M_Z^2) \;=\; 0.1191^{+0.0048}_{-0.0071} \tag{23}$$

While this value is less precise than those obtained in e^+e^- and other collisions, proton colliders measure the strong coupling at energy scales which have been out of reach up to now.

5.3.2 Fat Jets

Final state radiation also becomes visible in the particle and jet distributions inside a 'fat' jet, i.e. a jet with a large cone size. This concept has emerged as an important tool in searches for new particles that decay into multiple quarks or gluons [32]. Outstanding examples are Higgs bosons decaying into, e.g. a pair of $b\bar{b}$ quarks, or boosted top quarks decaying into three quarks (cp. Sect. 8.2.2). Schematically one can assign the following properties to such decays and Standard Model gluon emission:

- A Standard Model fat jet stems, e.g. from $q \rightarrow qg$ or $g \rightarrow gg$ splitting. The emitted partons tends to have rather asymmetric energies and define a plane in which the final particles are distributed.
- The products of a two – body decay have fairly equal energies and are inside a plane.
- A three body decay, like from a top quark, also has a fairly equal share of the energies but the particles do not lie within a plane.

Observables have been suggested to categorize fat jets along their planar structure and homogeneous energy distribution. They can be separated into those that aim at finding substructures, like searching for subjects with a smaller cone, or inclusive observables like:

- The jet mass $M^2 \;=\; \left(\sum_i E_i\right)^2 - \left(\sum_i \mathbf{p}_i\right)^2$
- The planar flow, which measures whether the energy is spread inside a plane given by

$$P \;=\; 4 \times \frac{det I_E}{Tr(I_E)^2} \;\text{ with }\; I_E^{kl} \;=\; \frac{1}{M}\sum_i \frac{1}{E_i} p_{i,k}\, p_{i,l} \tag{24}$$

  The sum goes over all particles and $k,\ l$ are the directions orthogonal to the jet axis and yields P is 0 if the particles are within a plane and 1 if the particles are on a circle.
- The Angularity, which parametrises the two dimensional width inside a jet

$$A \;=\; \frac{1}{M}\sum_i E_i \sin^a\theta_i [1 \;-\; \cos\theta_i]^{1-a} \;\sim\; \frac{1}{M}\sum_i E_i \,\frac{\theta^6}{8\theta^2} \tag{25}$$

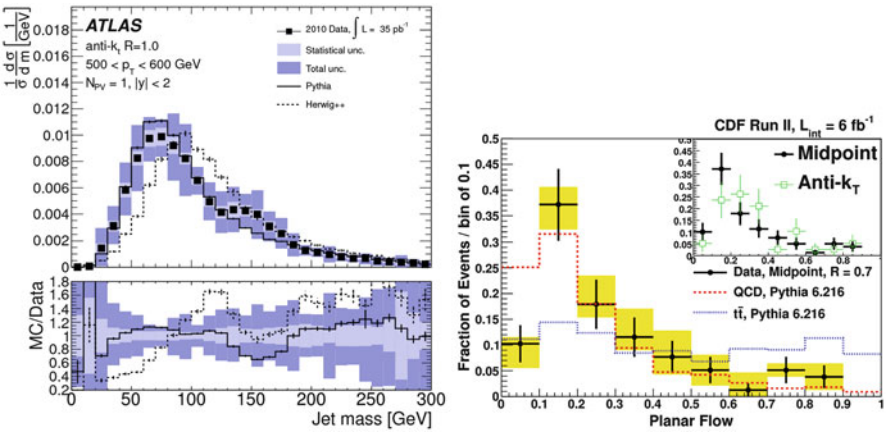

Fig. 10 Measurement of observables to parametrise the 'fat jet' structure [33]. *Left*: the jet mass. *Right*: the planar flow [34]

where θ_i is the angle of particle i relative to the jet axis. The second equation assumes $a = -2$, which is frequently used and highlights particles at high angles.

QCD predictions for the internal jet substructures have been tested in terms of these observables with a view to studying new heavy boosted particles. As examples, the mass for a jet p_T around 500 GeV at the LHC [33] and the planar flow at the Tevatron [34] are shown in Fig. 10. Whereas the Tevatron result is statistically limited, the jet mass shows that the PYTHIA expectation is too soft, and the HERWIG prediction too hard.

Different correction procedures for pile-up contributions than those discussed in Sect. 5.2 are used in fat jets. For example, by 'trimming' a fat jet, any subjet i is rejected that has a $p_{T,i} < f_{cut} \cdot p_{T,jet}$, where f_{cut} is a predefined value.

6 QCD and W and Z Boson Production

The colourless $W^{\pm}$ and Z^0 bosons offer rather clean QCD tests since they are less affected by colour flow between the proton remnants or final state radiation than di-jet production. The vector bosons at hadron colliders are produced by the Drell-Yan process in which a quark and an antiquark from the incoming hadrons merge into a γ, Z^0 or $W^{\pm}$. This process led to the milestone discoveries of the J/ψ, the Υ and the electroweak gauge bosons. Its cross section is a convolution of the $1/M^2$ dependence, typical for pointlike interactions and the parton luminosity and therefore steeply falling.

6.1 Production of Electroweak Bosons at the LHC

Because of the overwhelming jet production and the less precise measurements of their momenta, the standard way to study $W^{\pm}$ and Z^0 production at hadron colliders is to use their leptonic decay modes with branching ratios of [9]

$$\mathrm{BR}(Z^0 \to e^+e^-) \;=\; \mathrm{BR}(Z^0 \to \mu^+\mu^-) \;=\; 3.3658 \pm 0.0023\,\% \tag{26}$$

$$\mathrm{BR}(W^{\pm} \to e\nu_e) \;=\; \mathrm{BR}(W^{\pm} \to \mu\nu_\mu) \;=\; 10.80 \pm 0.99\,\% \tag{27}$$

As can be seen in Fig. 11 the Z^0 reconstruction yields an almost background free and narrow signal [35]. The approximately million Z^0 bosons per fb^{-1} produced at 8 TeV allow a broad physics program; in addition they are an important tool to calibrate the electromagnetic part of the detector. The cross section $\sigma(pp \to W) \cdot \mathrm{BR}(W \to l\nu)$ is even a factor 10 larger; however, the reconstruction of the $W^{\pm}$ bosons is complicated by their decay into neutrinos which can only be determined indirectly. Their momenta are approximated by the unobserved, 'missing' momenta obtained from requiring a momentum balance. Because of acceptance losses and large experimental uncertainties in the forward direction, the missing p_z component cannot be determined such that the missing momentum is restricted to the plane orthogonal to the beam direction. The $p_{T,miss}$ components are given by

$$p_{x,miss} \;=\; -\sum_i (p_x)_i \qquad p_{y,miss} \;=\; -\sum_i (p_y)_i \tag{28}$$

where the sums run over all measured objects, i.e. clusters in the calorimeter and muons.

Using $p_{T,miss}$ and the transverse momentum of the charged lepton, the transverse mass of the W boson decay can be measured and a clean signal extracted, as shown in Fig. 11 (right).

The cross section $\sigma \times \mathrm{BR}$ is shown in Fig. 12 for various energies and for NNLO predictions. Below a c.m. energy of 2 TeV measurements are due to $p\bar{p}$ collisions, and above 2 TeV from pp collisions. The cross section for $\bar{p}p$ is somewhat larger than the one for pp collisions because of the excess of valence antiquarks in antiprotons. On the other hand, in pp collisions antiquarks are only sea quarks. Furthermore there is an excess of the two valence up against the one down quark such that more W^+ than W^- are produced. The theoretical calculations for $\sigma \times \mathrm{BR}$ are in very good agreement with the data.

Moreover, since these valence quarks are at fairly high x, they in general lead to boosted W bosons. Therefore, the different production yields become more strongly visible at high $|\ \eta\ |$ and can be measured as an asymmetry between positive and negative leptons

$$A_l \;=\; \frac{N_{l+}(|\ \eta\ |) \;-\; N_{l-}(|\ \eta\ |)}{N_{l+}(|\ \eta\ |) \;+\; N_{l-}(|\ \eta\ |)}. \tag{29}$$

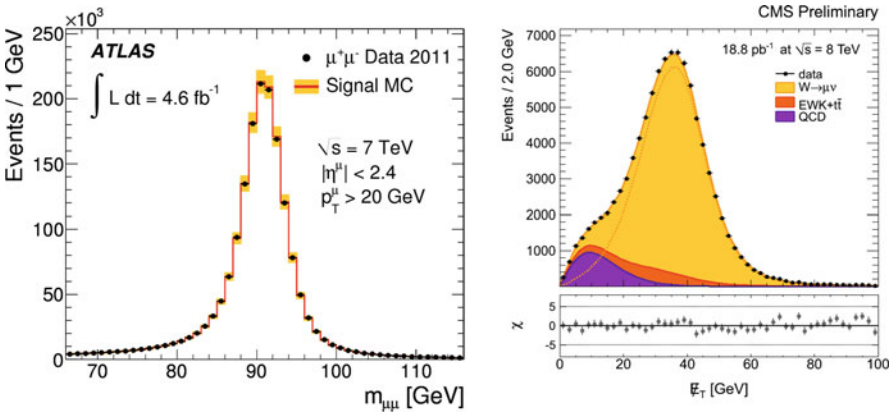

Fig. 11 Mass spectra of $\mu^+\mu^-$ [35] (*left*) and μ and missing transverse energy [36] (*right*) in events with isolated muons

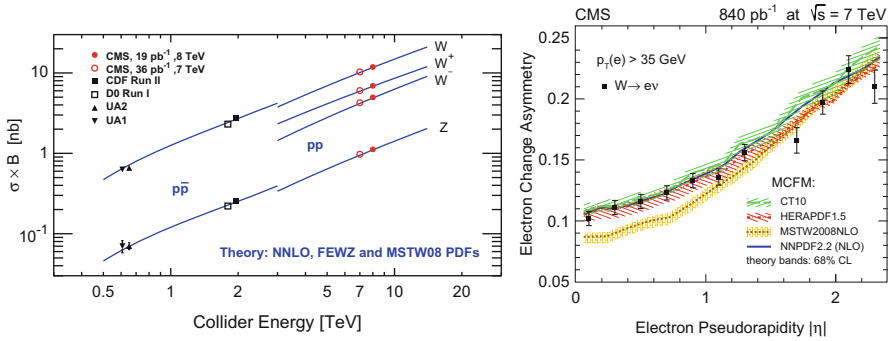

Fig. 12 Measurement of the cross section of $W^{\pm}$ and Z^0 production [36]. Electron charge asymmetry as a function of $|\eta|$ [37]

As can be seen in Fig. 12 (right) the asymmetry rises with the pseudorapidity. Whereas this trend is common to all parton distribution functions, their detailed magnitude and η dependence differ. Therefore, these measurements constrain pdfs. Additional constraints from LHC measurements are given by the ratio of inclusive Z^0 and $W^{\pm}$ production, respectively the W^+ and W^- cross sections. These have been discussed at this School.

As mentioned above, historically the Drell-Yan process was a discovery channel. The LHC extends its mass reach to several TeV. Results of such searches are shown in Fig. 13. Neither in $q\bar{q} \to e^+e^-$ nor in $q\bar{q} \to e\nu$ are deviations from the expected cross section observed, leading to lower mass limits on new Standard Model-like Z' or W' particles of 2.96 and 3.35 TeV respectively [38].

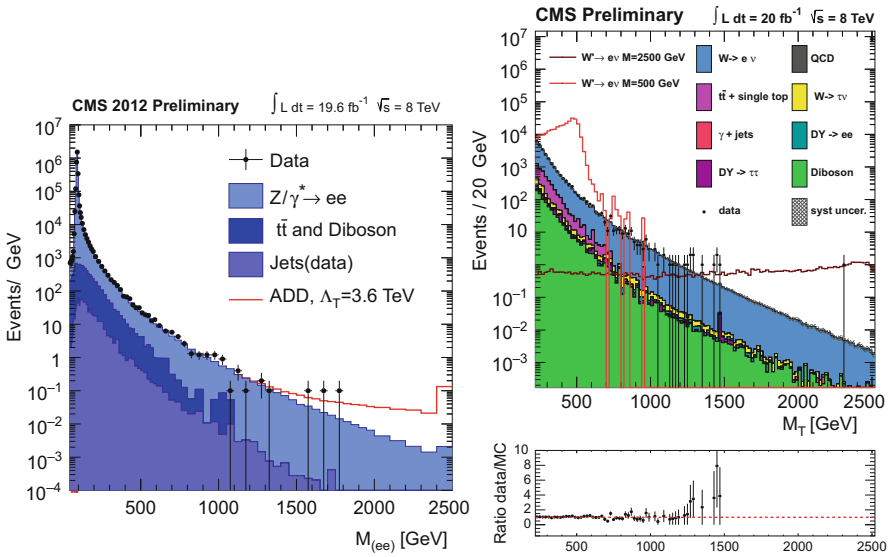

Fig. 13 Cross-section measurements in $\mu^+\mu^-$ and $\mu\nu$ production [38]

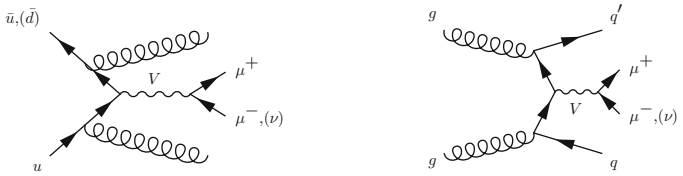

Fig. 14 Generic Feynman diagrams for Drell – Yan production in association with two jets

6.2 *Associated Jet Production*

There are several reasons to study W and Z bosons produced in association with jets, as depicted generically for two jets in Fig. 14:

- The relative simplicity of the process allows for rather precise and higher order QCD predictions.
- The associated production with charm and bottom quarks helps one to constrain the rather uncertain heavy quark content of protons.
- It is an important background for top quark production and many processes beyond the Standard Model.

A nice way to study these process is to determine ratios of processes, where many systematic uncertainties tend to cancel.

As proposed in [39], the fractional decrease in cross-section with the number of jets,

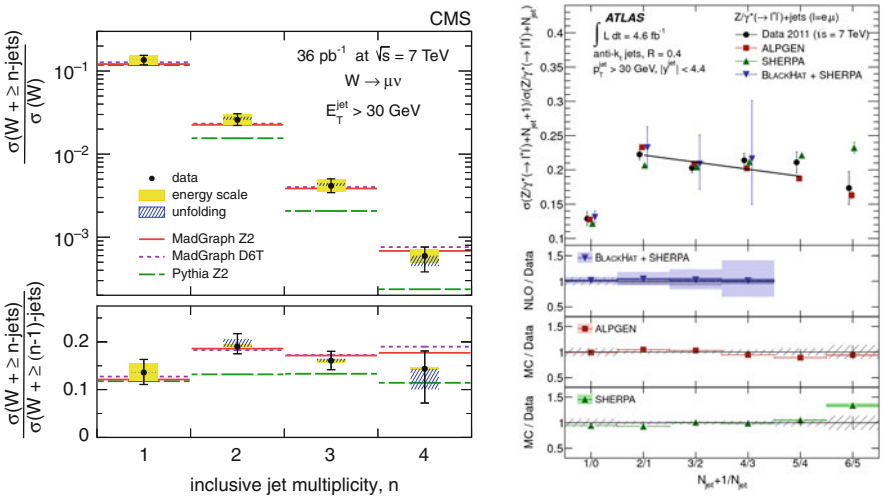

Fig. 15 Experimental tests of the Berends Giele scaling. Inclusive ratios of jet multiplicities in $W^{\pm}$ events [40] (*left*), differential ratios of jet multiplicities in Z^0 events [41] (*right*)

$$R(n) = \frac{\sigma(V + (n+1)\text{ jets})}{\sigma(V + n\text{ jets})} \sim \text{constant}, \tag{30}$$

should be rather uniform and of $\mathcal{O}(\alpha_s)$. This is called Berends-Giele scaling.

Figure 15 shows experimental tests of this relation based on $36\,\text{pb}^{-1}$ and $5\,\text{fb}^{-1}$ [40, 41]. Depicted on the right side are the ratios $R(n)$ for W and on the left for Z^0 bosons. Indeed, within the experimental uncertainties $R(n)$ is fairly constant with a value of 0.15–0.2 for both W and Z bosons. It should, however, be noted that $R(n)$ depends strongly on the requirement on the jet p_T. There seems to be some tension between PYTHIA and the data for the W sample, whereas the generators compared to the high statistics Z^0 data are in excellent agreement.

As also suggested by Fig. 14, parton production is rather similar in association with either vector boson. Therefore, Z and W + jet events should have fairly similar properties. The ratio of the jet p_T in $W, +1$ jet events has been measured [42]. Jets associated with W bosons tend to have a lower p_T, however, for $p_T > 100\,\text{GeV}$ the ratio between W/Z is leveling out.

Of special interest is the study of associated production with bottom jets. It has the potential of constraining the heavy quark distribution functions. Moreover, V+bottom production is an important background for top and the Higgs strahlung from W/Z.

Whereas the Zbb coupling is fairly large, the Wbc coupling is suppressed due to the smallness of the CKM matrix element $|\ V_{cb}\ |^2$. Therefore the fraction of bottom jets in Z^0 events is expected to be larger than in $W^{\pm}$ events. Measurements based on $2.1\,\text{fb}^{-1}$ show a fair agreement with the expectation for the cross section of single and double bottom production in association with the Z^0 [43]. Within the

experimental cuts a cross section of some 0.4 pb was measured for $Z^0 + b\bar{b}$, which has to be compared with the inclusive Z^0 cross section which is larger by a factor of around 2,500. Also the event properties agree with the expectation. Data also show the expected strong suppression of W + b jets production [44].

7 Electroweak Measurements

Electroweak precision measurements were the highlight of the 1990s especially at CERN's e^+e^- collider LEP. Notably, the determination of the Z boson mass was a masterpiece of experimental and theoretical research that accounted for moon tides, schedules of the Swiss railways, sub-per-mil level experimental acceptance determination, and high order corrections to the Z^0 resonance shape.

Proton collisions now provide the best measurements of the $W^{\pm}$ mass. They also contribute significantly to the tests of the electroweak theory by the precision measurement of the top – quark mass, and now finally adding the Higgs boson. Moreover, the Tevatron and LHC test electroweak theory directly at scales which are far beyond the energy reach of LEP.

At lowest order all electroweak measurements are unambiguously predicted by three parameters, which are conventionally chosen as [9]:

$$\alpha_{em}(0) = 1/137.03599976(50) \tag{31}$$

$$G_{\mu}(m_{\mu}^2) = 1.6639(1) \cdot 10^{-5} \ \text{GeV}^{-2} \tag{32}$$

$$M_Z = 91.1882(22) \ \text{GeV} \tag{33}$$

fixed at Q^2 of 0, the muon mass, and the Z^0 boson mass.

However, these parameters alone cannot consistently describe the electroweak measurements. Instead, Standard Model calculations have to account for quantum fluctuations, e.g. due to the top quark and the Higgs boson. During the last decades a central aim has been to reach highest sensitivity to their effects. High precision theoretical calculations and measurements have led to the fantastic achievements of determining the mass of first the top quark and finally of the Higgs Boson, from virtual effects before their direct observation.

7.1 Helicity Structures of Production and Decay

The parity-violating coupling of fermions to the W boson provides an interesting tool to determine its polarisation. The basic idea is sketched in Fig. 16. Assume a W^+ to be produced by a valence u quark and a $\bar{d}$ quark from the sea. Due to the weak V-A coupling the spin of the W^+ is opposite to the flight direction of

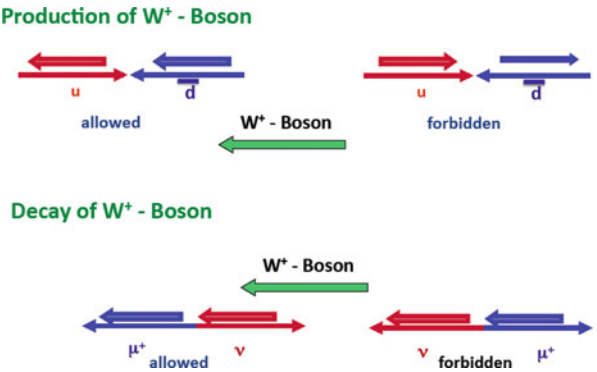

Fig. 16 Cartoon of the helicity structure of production and decay of a W^+. Spin directions are given by the thick, flight directions by *thin arrows*. The directions of production and decay are assumed to be aligned with the spin of the W boson

the up quark, and therefore has a negative helicity. Again the V-A coupling of the $W^+ \rightarrow \mu^+\nu$ dictates the muon to fly along the spin direction of the W.

This is an idealised picture. More precisely, the distributions of θ^*, the angle of the charged lepton wrt. the flight direction of the boson, are given for W's with negative (positive) helicity $W_{-,(+)}$ and longitudinal W_L's as

$$W_\pm \propto \frac{3}{8}\left(1 \mp \cos\theta^*\right)^2, \qquad W_L \propto \frac{3}{4}\sin^2\theta^* \tag{34}$$

As discussed in Sect. 6.1, at the LHC, W boson production depends on the rapidity. If the W^+ originates from a valence u-quark, which happens mostly at high $|\ y\ |$, its spin is aligned against the flight direction. In contrast, a W^+ produced at $y \sim 0$ is mostly due to up quarks from the sea and therefore does not have a preferred boost direction. In this case both helicity states are possible.

Whereas for W bosons produced collinearly to the beam direction, only left and right handed polarisations are possible, longitudinal W bosons can be produced if they have a transverse momentum. QCD calculations show, however, that for $p_T \rightarrow \infty$ most of the produced W bosons are left handed, some 20 % right handed, whereas the longitudinal polarisation vanishes [45]. In the general case the differential distribution $d^2\sigma/dp_T^W dy$ is more involved. The experimental selection, especially the requirements on a minimum p_T of the missing momentum and the charged lepton, significantly distorts the measured distribution and tends to wash out the difference between these polarisation states. Still, measurements have discriminating power, e.g. based on 31 pb^{-1}, good agreement of the data with the theoretical expectation was found [46].

Similarly, the polarisation of the Z^0 can be determined from the asymmetry of the leptonic decay products. This was a crucial measurement at LEP. Whereas the precision at the LHC is significantly less, it extends this measurement to masses

of several hundred GeV and therefore provides additional sensitivity to deviations from Standard Model expectations, e.g. due to a high Z' mass. Analyses at 7 TeV find a good agreement between data and theory [47].

7.2 Mass of the W Boson

The mass of the W boson is one of the fundamental predictions of the Standard Model once its basic parameters are fixed. In leading order it is given by the parameters of Eq. 31ff:

$$(M_W^2)_{LO} = \frac{M_Z^2}{2}\left(1 + \sqrt{1 - \frac{4A}{M_Z^2}}\right) \quad \text{with} \quad A = \frac{\pi\alpha}{\sqrt{2}G_\mu}. \tag{35}$$

Radiative corrections, denoted by Δr, whose sizes depend e.g. on the masses of the top quark and the Higgs boson, change the expression to

$$(M_W^2)_{rad} = (M_W^2)_{LO}\,\frac{1}{1-\Delta r}. \tag{36}$$

As illustrated in Fig. 18, measuring M_W fixes Δr and thus constrains M_{top} and M_{higgs}. To arrive at a significant result, however, the M_W has to be known to better than 0.1 %. This challenge has been met by LEP's measurement of $M_W = 80396 \pm 33$ MeV [48].

At proton colliders W bosons are identified by their decay into $l\nu$. However, as discussed in Sect. 6.1, instead of M_W only the transverse mass M_T is available

$$M_W^2 > M_T^2 = (E_{T,l} + E_{T,miss})^2 - (\mathbf{p}_{T,l} + \mathbf{p}_{T,miss})^2 \sim 2p_{T,l}\,p_{T,miss}\,(1-\cos\Delta\phi) \tag{37}$$

The relation between M_T and M_W can be used to precisely measure the W mass.

An alternative observable is the transverse momentum p_T of the charged lepton. In contrast with M_T, it avoids the somewhat uncertain $E_{T,miss}$, but is affected by a possible p_T of the W boson due to QCD effects. Neglecting these, the relation with M_W is given by

$$p_T(l) = \frac{M_W}{2}\,\sin\theta^* \tag{38}$$

where θ^* is the angle with respect to the beam in the W boson rest system. Therefore

$$max(p_T(l)) = \frac{M_W}{2}. \tag{39}$$

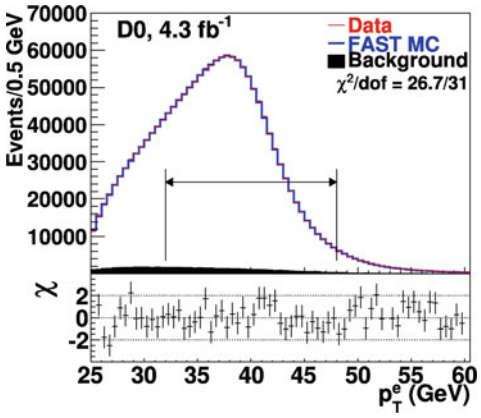

Fig. 17 The p_T of the electron in $p\bar{p} \rightarrow l\nu$ events [49]

For an isotropic decay, the most populated p_T on the surface on the sphere in momentum space of the decay products is $max(p_T(l))$. More formally,

$$\cos\theta^* = \sqrt{1 - 4 \cdot p_T^2/M_W^2} \tag{40}$$

and, with $d\sigma/d\cos\theta^*$ being the decay angular distribution,

$$\frac{d\sigma}{dp_T^2} = \frac{d\sigma}{d\cos\theta^*} \frac{2 \cdot M_W}{\sqrt{M_W^2 - 4 \cdot p_T^2}} \tag{41}$$

exhibiting a pole at $p_T = M_W/2$ which in reality is broadened by the W width $\Gamma_W = 2.085 \pm 0.042$ GeV.

Initial and final state radiation, the underlying event and experimental effects further diffuse this sharp fall off. By exploiting the similarity of W and Z^0 production and the superb precision of M_Z (cf. Sect. 7), the systematic uncertainties of the M_W measurements can be kept very small.

Recently, measurements of M_W by D0 [49] and CDF [50] have superseded LEP's precision. CDF uses as default M_T, while the D0 analysis prefers the electron p_T.

D0 first selects W boson candidates requiring $p_T < 10$ GeV for the residual event, thereby reducing QCD biases. The observed $p_T(e)$ distribution (see Fig. 17) is then compared with a well-tuned simulation. Varying the nominal mass of the W boson and comparing the likelihood to the data, the best agreement is found for $M_W = 80.342$ GeV. Systematic uncertainties are largely estimated from comparisons with Z^0 boson production:

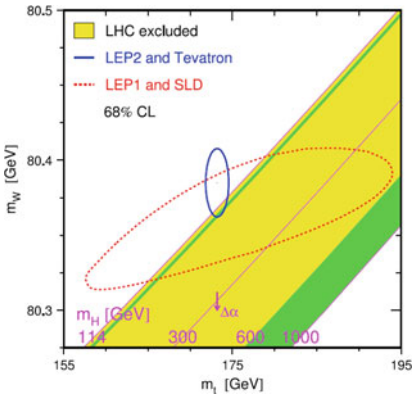

Fig. 18 Mass of W vs. mass of top and the regions for different Higgs masses [48]

- The energy scale of the electron could shift the apparent mass away from the true mass. Since the masses of the W and Z bosons are not too far apart, such a shift, however, should also lead to a wrongly measured Z^0 mass. Calibrating the electron scale to match the correct Z^0 mass also fixes the energy scale of the electron from the W boson decay.
- The energy resolution washes out the sharp drop – off and therefore could also imply a shift in the reconstructed mass. Again the precisely measured Z^0 line shape provides a handle to estimate the detector resolution. The measured Z^0 width is a convolution of the detector resolution and the natural width, which has been accurately measured as $2.4952 \pm 0.0023\,\text{GeV}$ [9]. Unfolding of the observed width allows the detector resolution to be disentangled.
- Finally also the smearing of the p_T distribution due to QCD effects can be estimated by comparing the p_T of an electron from W and from Z^0 decay, taking into account the different masses of the bosons and a small sensitivity to parton distribution functions.

The final combined statistical and systematic uncertainty yields 23 MeV. Combining this result with the CDF measurement, yielding $\sigma(M_W) = 19\,\text{MeV}$, and the one of LEP, the W mass is given as [51]

$$M_W = 80.385 \pm 0.015. \tag{42}$$

This 0.02 % precision is the fruit of 30 years of development of theoretical calculations and experimental methods. It shows the potential of hadron colliders to contribute significantly also to high precision physics and is a key result to constrain the mass of the Higgs boson from radiative corrections, Fig. 18.

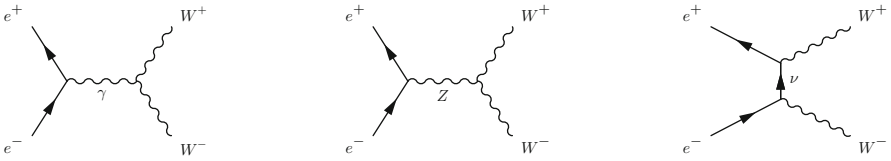

Fig. 19 Feynman diagrams of lowest order $e^+e^- \rightarrow W^+W^-$ production

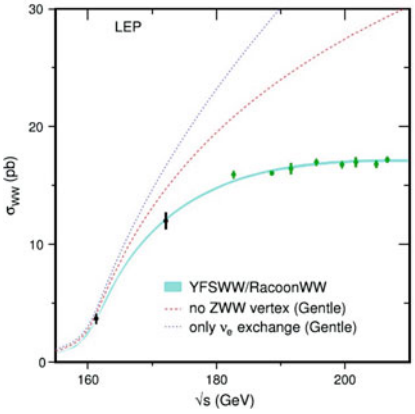

Fig. 20 Measurement of the cross section of $e^+e^- \rightarrow W^+W^-$ production at LEP [48]. Also shown are the theoretical expectations with the gauge cancellations of the Standard Model and alternatives if gauge contributions would be missing

7.3 Triple Boson Coupling

The self-coupling $(\gamma,\ Z^0)W^+W^-$ of the vector bosons is a central part of the electroweak SU(2) × U(1) gauge theory. Its strengths and dynamics address a core feature of gauge symmetries. Their relevance is underlined by historical precedents. It required a Z^0 to keep $\sigma(e^+e^- \rightarrow W^+W^-)$ finite at large $\sqrt{s}$ and also the Higgs boson is needed to regulate $W_LW_L \rightarrow W_LW_L$ scattering at energies of around 1.2 TeV. Moreover, it requires finely tuned compensations of the three boson coupling with the coupling of bosons to fermions to guarantee unitarity.This can e.g. best be seen in the process $e^+e^- \rightarrow W^+W^-$, whose three generic contributions are depicted in Fig. 19. Local gauge symmetry relates the couplings of $We\nu$ and those of WWZ, $WW\gamma$, and these couplings cancel each other such that the cross section does not violate unitarity.

LEP [48] has confirmed this prediction, as can be seen in Fig. 20. Whereas the data are in agreement with the Standard Model, a ν exchange contribution alone would blow up the cross section, as would a contribution from only ν and photon exchange.

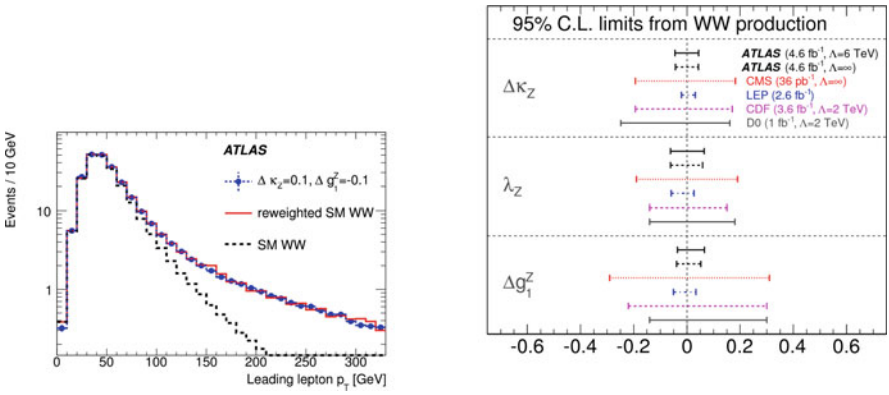

Fig. 21 ATLAS simulation of the leading p_T distribution of a lepton in W^+W^- events [53]. Shown are the Standard Model expectations together with those for anomalous triple gauge couplings (*left*) and limits on anomalous triple gauge couplings from ATLAS compared to other measurements (*right*)

The LHC has the potential to probe the triple boson vertex at energies of several hundred GeV, far beyond the LEP reach, and allows separation of the Z^0 and γ couplings through the processes $W \rightarrow WZ$ and $W \rightarrow W\gamma$, which was hardly possible at LEP. The WWZ interactions can be described by an effective Lagrangian (see, e.g. [52])

$$\frac{\mathscr{L}_{WWZ}}{g_{WWZ}} = i\left[g_1^Z\left(W^\dagger_{\mu\nu}W^\mu Z^\nu - W_{\mu\nu}W^{\dagger\mu}Z^\nu\right) + \kappa^Z W^\dagger_\mu W^\nu Z^{\mu\nu} + \frac{\lambda}{m_W^2}W^\dagger_{\rho\nu}W^\mu_\nu Z^{\nu\rho}\right] \tag{43}$$

and correspondingly for $WW\gamma$. Here $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$, and the overall couplings g_{WWZ} ($g_{WW\gamma}$) are $-e\cot\theta_w$, ($-e$). In the Standard Model, κ and g_1 are equal to 1 and $\lambda = 0$, and measurements are given in terms of deviations from these values. These couplings have different helicity structures, e.g. g_1 has a strong sensitivity to W_LW_L scattering, which in turn means that potential deviations have different energy dependencies, i.e. Δg_1 and $\Delta\lambda$ would grow with $\hat{s}$, the mass of the hard scatter. whereas κ grows with $\sqrt{\hat{s}}$. Thus the high energies of the LHC offer particular sensitivity to g_1 and λ.

As an example, the $pp \rightarrow W^+W^-$ process at LHC [53] is selected by requiring two oppositely charged isolated leptons and substantial missing transverse energy. Background due to the Drell – Yan process $pp \rightarrow \gamma, Z^0 \rightarrow l^+l^-$ or top quark events is suppressed by requiring the missing momentum to have a large angle wrt. to one of the leptons and by vetoing events with a jet. Since a larger lepton p_T is correlated with higher M_{WW}, deviations from the Standard Model become more prominent for high p_T leptons as schematically shown in Fig. 21 (left).

The data are in agreement with the Standard Model expectations, and exclude large values of anomalous couplings. In Fig. 21 (right) the ATLAS limits, obtained with $4.6\,\text{fb}^{-1}$ at 7 TeV, are compared to those of LEP and other experiments. Whereas the LEP precision on $\Delta\kappa$ is still a factor 2 better than those at the LHC, those for λ_Z and Δg_1^Z are comparable. These early results show the potential of the high energy proton collisions.

8 The Top Quark at Hadron Colliders

8.1 Basic Properties of the Top Quark

8.1.1 Search and Observation

The top quark seems to be special. Not only is its mass of $173.18 \pm 0.94\,\text{GeV}$ [54] the heaviest Standard Model particle, but its Yukawa coupling to the Higgs boson

$$\lambda_t \;=\; \sqrt{2}\,\frac{M_t}{\text{v}} \;=\; 0.996 \pm 0.006 \tag{44}$$

is consistent with unity. Here $\text{v} = 246\,\text{GeV}$ is the vacuum expectation value. Of course, a value $\lambda_t = 1$ can be purely accidental, but it is certainly intriguing and raises questions about a possible special role of the top quark in electroweak symmetry breaking. Some theorists conclude that it is the only fermion with a 'natural' mass.

The search for the top quark started in 1973 after the τ lepton as first element of the third generation was found. At that time nobody expected it to be so heavy. Instead physicists were sure to find it at a mass of around 15 GeV, being fooled by the apparent regularity of the masses of $J^P = 1^- q\bar{q}$ states:

$$M(s\bar{s}) = 1.02\,\text{GeV}, \qquad M(c\bar{c}) = 3.1\,\text{GeV}, \qquad M(b\bar{b}) = 9.4\,\text{GeV},$$

Surviving the searches of many accelerators, whose strong motivation was to find the top quark (PETRA/PEP, TRISTAN, LEP, S$p\bar{p}$S), it was finally observed in 1995 at the Tevatron. At that time, its mass was no longer a surprise, as electroweak precision data and theory had already constrained it to $M_t = 178.8 \pm 20\,\text{GeV}$. The observation at a mass consistent with expectations from quantum fluctuations, is one of the astounding successes of the Standard Model. Given this high mass it can, as yet, only be explored at proton colliders.

Table 2 Measurements of top properties in its different top decay channels

Observable	$t \to blv$	$t \to bq\bar{q}$
Charge sign	Yes	Difficult
Momentum	With constraints	Yes
Helicity	Yes	No
Mass	With constraint	Yes

8.1.2 General Expectations

Apart from the precise determination of its mass, experiments probe whether top properties are like those of other Standard Model fermions. The Tevatron pioneered top physics with up to a few thousand events. At the LHC now, up to around 50,000 $t\bar{t}$ events are produced per day. After experimental selections, some 400,000 $t\bar{t}$ pairs can currently be used.

The Standard Model makes unambiguous predictions for the top quark. A unique feature is that it decays weakly before strong interactions can form a bound state. No top – hadrons exist, and the top quark can be regarded as (almost) a free quark. Furthermore, unitarity of the CKM matrix imposes that 99.8 % of the top quarks should decay into a bottom quark. These clearly defined Standard Model properties lead to a well defined experimental check list for probing the top quark.

The almost exclusive decay $t \to bW^+$ leads to three kinds of signatures for $t\bar{t}$ production with relative yields reflecting the branching fractions of the W boson.

1. $t\bar{t} \to 6$ jets, i.e. where both W bosons decay into quarks. This has the largest fraction of 0.46, however, suffers from considerable background.
2. $t\bar{t} \to l\nu +$ 4 jets, i.e where one W boson decays into quarks and the other leptonically. Its fraction is 0.44, almost as frequent as the fully hadronic decay. However, since the τ lepton reconstruction is less efficient and its analyses more complicated, one mostly uses just the decays into e and μ, reducing the fraction to 0.29. The background, mainly due to W + jets events, is relatively small.
3. $t\bar{t} \to l^-\bar{\nu}l^+\nu +$ 2 jets has a fraction of 0.10 for just e and μ decays. This decay is largely background free.

In all channels two of the jets are due to bottom quarks. Utilising the highly efficient bottom identification, background can be reduced further.

Beyond the yields alone, these channels hold different physics information as summarised in Table 2. In general the semileptonic decay channel, with its fairly high yield, its relatively low background and its good physics potential is the favoured one. On the other hand, most analyses can be performed in all of the three $t\bar{t}$ signatures, albeit with significant differences in precision.

The experimental program on top quarks in the $t\bar{t}$ channel can be separated into tests of production properties, probing QCD and the gtt vertex; and decay properties, testing the weak coupling of the top quark. In addition the weak properties can be tested with single top production.

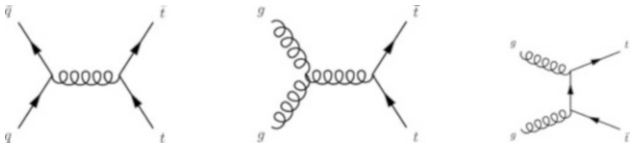

Fig. 22 Feynman diagrams for tt production

8.1.3 Production of Top Quark Pairs

The required energy fractions x of the partons to produce the $t\bar{t}$ pair are given by

$$\sqrt{x_1 \cdot x_2} \geq \frac{2 \cdot M_t}{E_{pp}} \tag{45}$$

which translates into a typical $x_{Tevatron} \sim 0.18$ and $x_{LHC} \sim 0.05$ at 7 TeV running, and 0.025 for the future 14 TeV LHC running. Thus, at the Tevatron $t\bar{t}$ pairs are predominantly quark induced but gluon induced at the LHC. The generic Feynman diagrams for top – pair production are shown in Fig. 22.

8.2 Measurements on $\mathrm{t\bar{t}}$ *Production*

8.2.1 Cross Section of $\mathrm{t\bar{t}}$ Production

Although the calculation of the cross section of the massive top quarks faces considerable technical challenges, quite precise results exist. The measurement of the cross section is therefore an important QCD test. In a first step, competing background processes are suppressed and a fairly clean sample of top – pair events are obtained. Typically, the semileptonic $t\bar{t}$ channel is selected based on:

- Four jets of high p_T of a minimum (e.g. ∼40–25 GeV),
- An isolated electron or muon of high p_T (e.g. >20 GeV),
- Substantial missing transverse momentum (of >20 GeV),
- Often at least one of the jets should be bottom tagged.

With these requirements $N_{measured}$ events are retained and the cross section is given by

$$\sigma_{t\bar{t}} = \frac{N_{measured} - N_{background}}{\epsilon \cdot \mathscr{L}} \tag{46}$$

Here $N_{background}$ is the number of background events, ϵ the selection efficiency and $\mathscr{L}$ the luminosity. The key experimental task is to keep these three contributions under control as tightly as possible. The luminosity measurement is based on special

LHC runs that scan the beam profile which, together with the known beam intensity, leads to the luminosity. These runs are used to calibrate the special luminosity components in ATLAS and CMS. The current uncertainty is about 2–4 %.

Also for the other two contributions to the cross section method, a main goal is to determine their uncertainty with only minimal reliance on simulation.

On the other hand, the efficiency, especially if regions outside of the fiducial volume are used, has to be based on Monte Carlo generators. To this end NLO generators of $t\bar{t}$ production are used, with detector simulation, and selection cuts applied. Theoretical uncertainties are for example due to the implementation of parton showers to simulate the effects of higher order corrections. Detector uncertainties, like the jet energy scale, efficiencies for bottom tagging or lepton identification, are mostly determined from data. For example, the electron efficiency is obtained from a tag-and-probe method of $Z^0 \to e^+e^-$ events which starts with a Z^0 selection using wide cuts. The fraction of electrons retained by the final hard cuts yields the efficiency.

The other major experimental input to determine the cross section is the background. Here 'data driven' methods are frequently applied. In these methods a region B is selected which uses similar but not identical requirements and object definitions as the signal region S, such that the background is dominant and the $t\bar{t}$ fraction is low. This region yields the normalisation of the background. Assuming the shape of the background simulation to be correct, the number of background events in B is extrapolated into S yielding the background contribution in S.

The main background to the $t\bar{t}$ process comes from W production associated with four jets (cf. Sect. 6.2). In a data-driven method, ATLAS estimates its contribution from the excess of W^+ over W^- discussed in 6.1. Such an imbalance does not occur in $t\bar{t}$ events and taking $r_{MC} = N_{W+}/N_{W-}$ from simulation, the number of W + jet events in the data is given by

$$N_W = (N_{W+} + N_{W-})^{data} = \frac{r_{MC}+1}{r_{MC}-1}(N_{W+} - N_{W-})^{data} \qquad (47)$$

Since the statistics are not sufficient for the important background of $W+$ b + 3 jets, the method is applied to events with 2 or 3 jets without b-tag requirement. Simulation is then used to extrapolate the result to the $t\bar{t}$ selection.

The top pair cross section has been determined in several tens of analyses using all decay channels and applying different methods. The LHC measurements yield [55]

$$7\ \text{TeV}: \quad 173.3 \pm 2.3 \pm 7.6 \pm 6.3 \ \text{pb}, \qquad 8\ \text{TeV}: \quad 239 \pm 2 \pm 11 \pm 6 \ \text{pb}. \qquad (48)$$

The 7 TeV result is a combination of ATLAS and CMS measurements; the 8 TeV is just from CMS. The uncertainties relate to the statistical, experimental systematic error and the one for the luminosity. The measurements agree with a theoretical prediction at approximate NNLO level of 167^{+17}_{-18} pb (7 TeV) and 238^{+22}_{-24} pb (8 TeV),

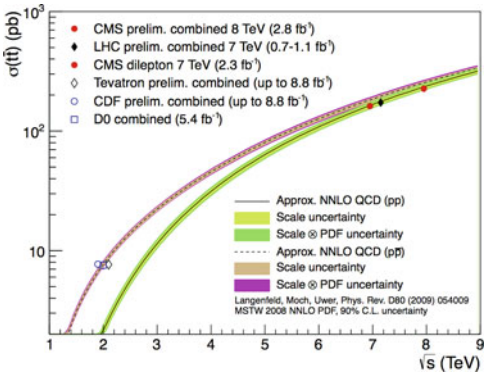

Fig. 23 Cross section of $t\bar{t}$ production at various pp ($p\bar{p}$) energies. The results are compared to approximate NNLO calculations [59]

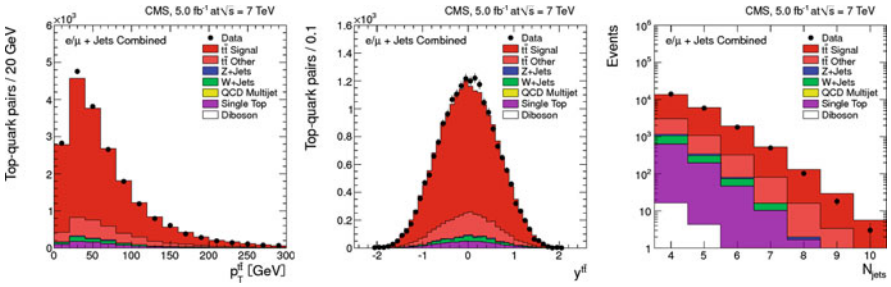

Fig. 24 Measurements of the p_T of the $t\bar{t}$ system, its rapidity, and the number of jets in $t\bar{t}$ events [60]

where the top mass is assumed to be 172.5 GeV [56]. Other theory groups predict largely the same cross section [57].[2] These results, together with those at the Tevatron, are shown in Fig. 23.

8.2.2 Differential Distributions of $t\bar{t}$ Pairs

QCD effects appear more significantly in differential distributions of top pair production. As examples, the p_T and rapidity distributions [60] are shown in Fig. 24 (left and centre). As yet all agree with the expectations from NLO calculations. Also the number of jets in a top pair event is well described, except, maybe, a small deficiency at multiplicities $\geq$6.

[2]Recently the cross section has been determined in full NNLO QCD corrections with NNLL soft gluon summation [58]. They obtain an improved precision of $172^{+6}_{-7.5}$ pb, and $245.8^{+8.8}_{-10.6}$ pb respectively.

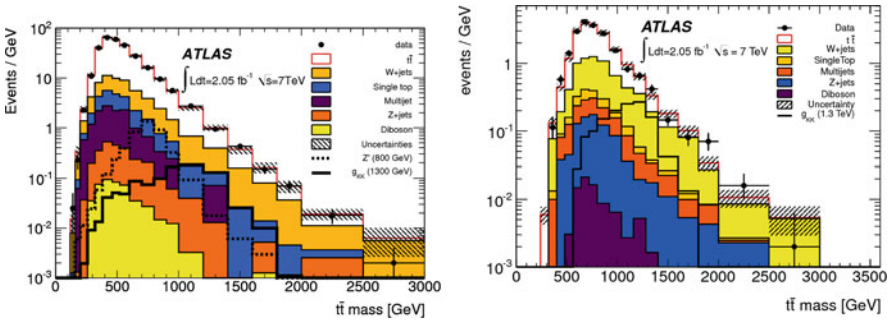

Fig. 25 Mass of the $t\bar{t}$ system using four jets (*left*) [62] and the fat jet method (*right*) [63]

The high LHC luminosity also allows first measurements of $t\bar{t}$ production associated with photons, W or Z^0 bosons. In the long term these will allow their couplings to top quarks to be measured. They are also an important background for searches and may be due to $t(W^{\pm},\ Z^0)$ resonances, postulated in several extensions of the Standard Model. The current measurements are in agreement with the expectations, but still have rather low statistics [61].

The mass distribution of the $t\bar{t}$ system is interesting in many respects. It is rather precisely predicted by QCD. At masses above 1 TeV it requires new methods for selection and reconstruction. At even higher masses, electroweak collinear corrections become important; and finally, quite a few BSM models predict resonances decaying into a pair of top quarks.

The $M_{t\bar{t}}$ distribution, obtained by ATLAS using the default selection discussed above, is shown in Fig. 25 (left). Here all four jets from top decays are resolved. Top pair masses of close to 3 TeV are probed. The data agree well with the prediction.

On the right side of the figure the mass spectrum is shown for a complementary selection, particularly suited for high $M_{t\bar{t}}$ where top quarks are boosted [63]. Their signature is different from the resolved one since the three jets from hadronic top decays merge. Therefore, a 'fat jet' instead of three narrow individual jets is required (cf. Sect. 5.3). Such a jet is selected requiring a large ΔR (e.g. 1) and a high mass (>100 GeV) and p_T (>250 GeV). In the next step, sub-jets inside the fat jet are required. As can be seen from the Figure, these fat jets add a significant number of events at high masses. Again, their mass distribution agrees with the QCD expectation.

Heavy particles which decay into a pair of top quarks have been postulated in many extensions of the Standard Model, e.g. extended Higgs sectors, Technicolour, extra dimensions or Little Higgs models (for a summary see e.g. [64]). The agreement of data with the QCD prediction means that a large range of masses and couplings of these hypothetical particles can be excluded. For example, at 95 % confidence the mass of a Kaluza-Klein excitation as assumed in models with additional spatial dimensions has to be larger than 1.9 TeV [65].

8.3 Measurements of the Top Mass

The importance of the mass of the top quark for electroweak precision physics has already been discussed in Sects. 7.2, and 8.1.2. It is the first time that the mass of a quark can be measured directly and need not be inferred from hadron masses. All decay channels and several methods have been used to measure the top quark mass from its decay products. Excellent precision has already been reached at the Tevatron.

8.3.1 Mass of the Top Quark from Decay Products

The most precise measurements of M_t comes from the semileptonic $t\bar{t}$ decay channel. Their main challenges are threefold:

- As for the measurement of M_W (Sect. 6.1), the missing momentum is used as a proxy for the neutrino. Once M_W is known, the p_z component can be inferred from the quadratic equations

$$M_W^2 = (E(l)+E(\nu))^2 - (p_x(l)+p_x(\nu))^2 - (p_y(l)+p_y(\nu))^2 - (p_z(l)+p_z(\nu))^2 \tag{49}$$

  with

$$E(\nu) = \sqrt{p_x^2(\nu) + p_y^2(\nu) + p_z^2(\nu)} \tag{50}$$

  This quadratic equation has two solutions for the only unknown parameter $p_z(\nu)$. According to simulation, in 70 % of all cases the solution with the smaller $\mid p_z \mid$ is the correct one, making this the preferred choice.
- Another ambiguity is how to assign the four jets to a specific top quark. Four combinations are possible:

$$j_A, j_B, j_C/j_D; \;\; j_A, j_B, j_D/j_C; \ldots.$$

  where the first three jets should come from the hadronically decaying top quark, the remaining jet should be merged with the leptonically decaying W boson. The combinatorics is reduced to two if both b-jets are identified.

  Again, constraints help in this case. Since the mass of the top and anti – top should be identical,[3] it follows that $M(jjj) = M(jl\nu)$. Furthermore, two of the jets should come from the decay of the W boson, which implies that one combination should yield $M(jj) = M_W$. The constraints lead to preferred combinations.

[3] Measurements support this hypothesis.

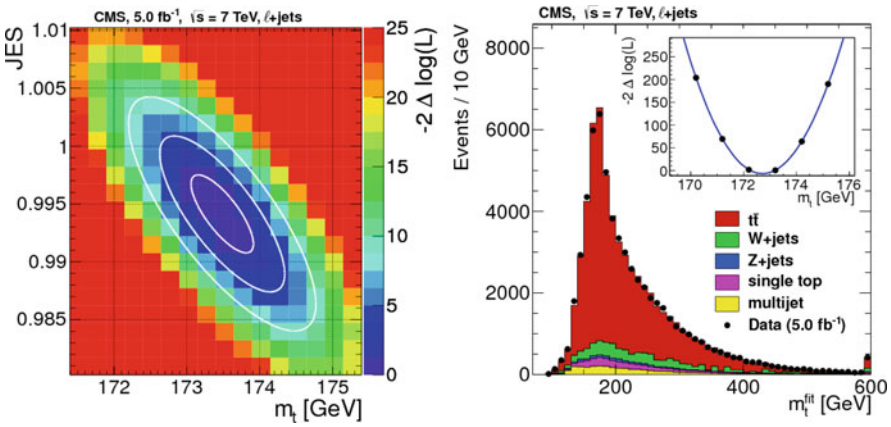

Fig. 26 Measurements of the top mass [66]. *Left*: correlation of the rescaling factor for the energy scale and the derived top mass. *Right*: mass distribution from four jets, the charged lepton and the neutrino from a kinematic fit and simulation with $M_{top} = 172.5$ GeV

- As for M_W (cf. Sect. 7.2) the precision of M_{top} depends directly on the uncertainty in the energy scale. Top decays provide a 'self – calibration' mechanism. The mass of the two-jet system from the W decay has to equal M_W. The jet energy scale, as determined by other means (see 5.2), can be modified to accommodate this. The amount of rescaling and the impact on the derived top mass becomes visible in Fig. 26 (left), which is the result of a two-parameter fit using the above constraints and leaving the top mass and the jet energy scale for the W boson mass free. As can be seen, the latter should be shifted up by 0.4 %. This scaling factor cannot be directly applied to the third bottom jet, which has a different particle composition, in particular contains rather frequently neutrinos.

Many methods are invoked to determine the top mass. Some use templates of simulated events assuming a specific top mass and determine for which mass they are in best agreement with the data. Another method is based on matrix element calculations of $t\bar{t}$ production and decay. Several 'true' top masses are assumed. For each M^i_{true} the multi-dimensional matrix element is integrated to assign a probability to each data event to agree with M^i_{true}. All effects mentioned above are taken into account, as well as experimental biases. The product of probabilities of all events leads to a likelihood value as a measure of the agreement of the data with M^i_t. The maximum of these likelihood values as a function of the true masses leads to the experimentally favoured top mass. Determining simultaneously the jet energy scale from the M_W constraint and the top mass [66], CMS obtains $173.49 \pm 0.43 \pm 0.98$ GeV, which is consistent with and almost as precise as the average of all Tevatron measurements of 173.18 ± 0.94 GeV [67].

The largest experimental systematic uncertainties are due to the jet energy scale correction for bottom jets (0.61 GeV). However there is also an uncertainty intrinsic to this method due to colour reconnection effects. The top – antitop system is

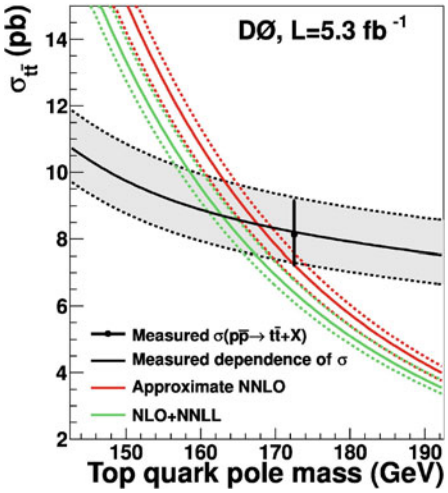

Fig. 27 Theoretical and measured cross section of $t\bar{t}$ production as a function of the top mass [71]

coloured and the top quark may interact softly with the coloured environment inside the proton. The colour flow in $t\bar{t}$ events can only be modelled [68] and not be calculated from first principles. Comparing different models of colour reconnection suggests an uncertainty of $\sim$0.5 GeV.

8.3.2 Top Mass Determination from Cross Section

As mentioned in Sect. 7.2, the top mass is a crucial ingredient to electroweak precision tests. The theoretical calculations require a mass in a well defined renormalisation scheme, mostly the 'pole mass'. The methods discussed in the previous subsection are based on templates which are constructed from QCD generators like PYTHIA [11], MC*at*NLO [69] or ALPGEN [70] using QCD corrections without a well defined scheme and it is unclear how these results fit into the electroweak precision calculations. It is folklore that it should be not differ by more than $\mathcal{O}(1\,\text{GeV})$ from the pole mass.

A better defined theoretical procedure is used for the calculation of the $t\bar{t}$ cross section which depends on the top mass. Therefore, the measurement of $\sigma_{t\bar{t}}$ (see Sect. 8.2.1) offers a complementary way to determine M_t. Figure 27 shows that the theoretical cross section depends strongly on the top mass. The prediction can be compared to the measured $\sigma(t\bar{t})$. Its slight mass dependence is due to selection efficiencies. The drawback of the method, however, are the rather sizeable theoretical and experimental uncertainties. Using the calculation of [72] D0 finds [73]

$$M_t^{\text{pole}} = 167^{+5.2}_{-4.7}\ \text{GeV}. \tag{51}$$

This value is consistent with the result from the direct measurement, although somewhat smaller.[4] Similar results are obtained from the cross sections measured by ATLAS and CMS [75].

8.4 Top Mass and Physics at the Planck Scale

The as-yet perfect agreement of data with Standard Model expectations means that they may be extrapolated to the highest mass scales without any theoretical problem. Then the first scale where new physics is known to exist is the Planck Scale, $M_P = 10^{18}$ GeV, where gravity becomes strong. *Gedankenexperiments* extrapolating the Standard Model up to this scale have some widely discussed implications.

- Firstly, it requires fine tuning of the renormalisation of the Higgs mass to 29 digits to arrive at the measured mass. This is to cancel loop contributions from the Standard Model particles, where, due to its large Yukawa coupling, the bulk of the one – loop corrections are due to the top loop (see, e.g. [76]). Whereas this general procedure of renormalisation is a standard, the hugeness of the correction, caused by the scalar and elementary nature of the Higgs boson is felt to be 'unnatural' and is one of the main driving forces to motivate extensions of the Standard Model. New symmetries are postulated that compensate the loop contributions, the most popular one being supersymmetry. This, for example, makes today's search for a light supersymmetric partner of the top highly interesting.
- The stability of the Higgs potential is characterized by parameters μ and λ. The value of μ is given by the W and Z boson masses, and the measurement of the Higgs mass of about 125 GeV fixes $\lambda(M_t) \sim 0.125$. Loop corrections, mainly due to the strong top coupling and the Higgs self interactions, let $\lambda(Q^2)$ decrease with larger Q^2. This is no problem as long as $\lambda > 0$. However, if λ turns negative, the Higgs potential exhibits a local minimum, the vacuum would become unstable and mass generation would be switched off. Extrapolating λ to the Planck scale and using the measured top mass, the universe is in a meta-stable state: if the top mass were 3 GeV lighter, the universe would be stable, but if it were around 5 GeV heavier the universe would be unstable, making us wonder why there are massive particles (and ourselves) in the universe (for recent discussions see, e.g. [77]).

These are inspiring considerations. New effects may be seen at LHC that solve the fine tuning problem. This in turn would also change the view of the stability at the Planck Scale.

[4] A complementary method has recently been suggested [74].

Fig. 28 Cartoon showing the helicity structure of top decays

8.5 The Helicity of Top Quarks in Hadron Collisions

The rapid weak decay of the top quark preventing hadron formation makes it possible for the first time to determine the helicity structure of a quark. A cartoon of the V-A helicity structure for a top decay is shown in Fig. 28. As discussed in Sect. 7.1, the W boson transfers its polarisation state to the decay angle of the charged lepton.[5]

8.5.1 The Helicity of Top Quark Decays

Defining θ^* as the angle of the charged lepton relative to the reversed momentum of the b-jet in the W rest system, the angular distribution can be written as

$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\theta^*} = \frac{3}{4}(1-\cos^2\theta^*)\,F_0 + \frac{3}{8}(1-\cos\theta^*)^2\,F_L + \frac{3}{8}(1+\cos\theta^*)^2\,F_R \quad (52)$$

where the F_i denote the longitudinal, left and right handed polarisation state of the W boson. As can be seen from Fig. 28 no right handed W bosons are allowed and therefore the forward direction, $\cos\theta^* > 0$, should be depleted. The data [78] are in agreement with the expectation (Fig. 29) and a fit to the measurement yields

$$F_0 = 0.67 \pm 0.03 \pm 0.06 \quad F_L = 0.32 \pm 0.02 \pm 0.03 \quad F_R = 0.01 \pm 0.01 \pm 0.04 \quad (53)$$

These results can be used to set limits on higher dimension operators due to Beyond the Standard Model contributions, affecting the left and right handed coupling of the top quark.

8.5.2 Top Spin Correlations

Whereas individual top quarks should not be polarised, the helicities of t and $\bar{t}$ should be correlated. The amount, sign of the correlations, and axis of alignment

[5] In principle also quarks can be used. As long as a fermion from the W decay can be distinguished from its anti-fermion, the W boson helicity can be measured. This is very challenging and only possible for a few quark species. Methods exist to identify e.g. the charge of a charm quark jet; however, these are rather inefficient.

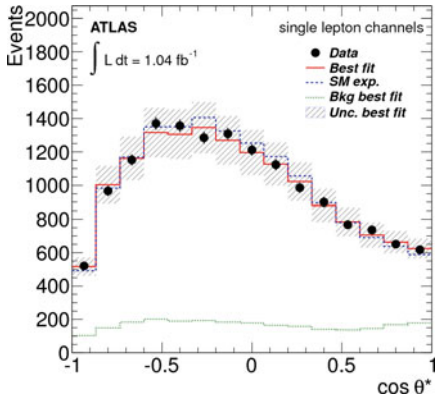

Fig. 29 Decay angle distribution ins the W rest system in top decays [78]. Also shown are the Standard Model expectations for a V-A coupling

depend on the production mechanism, by either $q\bar{q}$ or gg; and on $M_{t\bar{t}}$. They are predicted by QCD and, along the lines of the previous subsection, can be measured from from correlations between the lepton decay angles.

- The helicity structure of $q\bar{q} \rightarrow g \rightarrow t\bar{t}$ production is identical to the textbook example of $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$. Due to helicity conservation for a vector coupling, the spins of the initial quarks are aligned along their flight directions and at threshold, where the angular momentum of the $t\bar{t}$ system has to be $l = 0$, it is in a $^3S_1$ state. Angular momentum conservation means the spin directions of the $t\bar{t}$ are aligned with those of the incoming light quarks. For very high $\sqrt{\hat{s}} \gg 2 \cdot m_t$ the two top quarks behave like massless fermions and due to helicity conservation they have opposite helicities.
- The case of gluon scattering $gg \rightarrow t\bar{t}$ is more complicated. The incoming gluons can have both same and opposite helicities. It turns out that at threshold dominantly same helicities are produced, at higher energies unlike helicities. Thus, for small $m_{t\bar{t}}$ the $t\bar{t}$ system is in a $^1S_0$ state and also the top quarks have the same helicities. For high $t\bar{t}$ masses top quarks are relativistic, and helicity conservation requires them to have opposite helicities.

In all cases the spins of the outgoing top quarks are correlated, albeit that the axis, along which the correlation becomes maximal, depends on the production mode. The correlation can be observed using $\Delta\phi$, i.e. the difference in the azimuthal angles of the two charged leptons. Leptons that fly in the same direction, i.e. $\Delta\phi \sim 0$, tend to come from combinations of opposite helicities. Those that fly in opposite direction, i.e. $\Delta\phi \sim \pi$ tend to come from opposite spin directions, i.e. the same helicities. The ATLAS measurement [79] based on 2.1 fb^{-1} is displayed in Fig. 30.

Theoretical calculations have been performed assuming different quantisation axes yielding different magnitudes of the correlation. The ATLAS analysis compares the measurement to the prediction in the helicity basis along the flight

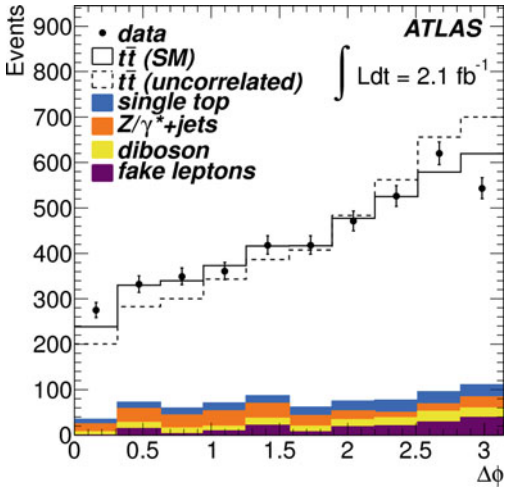

Fig. 30 Angle $\Delta\Phi$ between the charged leptons from a dileptonically decaying $t\bar{t}$ pair in the plane transverse to the beam direction. The measured distribution is shown as *dots* with error bars. Also shown is the Standard Model expectation (*full line*) and the expected distribution in the absence of spin correlations (Taken from [79])

direction of the top quarks. Taking into account experimental distortions, especially the requirement on a minimum p_T of the lepton, it yields

$$A_{hel} = 0.40 \pm 0.04^{+0.08}_{-0.07}, \tag{54}$$

consistent with the theoretical expectation of 0.31. In the long term it will be interesting to study these correlations as a function of $M_{t\bar{t}}$. In particular, spin correlations have been suggested to be a sensitive probe e.g. for scalar resonances decaying into $t\bar{t}$ (see Sect. 8.2.2), which would render like sign helicities.

8.6 Single Top Production

Whereas top quark pairs are produced in strong interactions, single top quarks can also be weakly produced at hadron colliders via a W boson interaction. Generic Feynman diagrams are depicted in Fig. 31.

Single top production opens a special window to the weak interactions of the top quark, e.g. it allows one to determine the CKM matrix element $|\ V_{tb}\ |$ and not only the branching ratio, as in top quark pair production. It is also sensitive to new particles, e.g. decaying into tb and to additional couplings.

At the LHC the dominant production channel is expected to be the t-channel W boson exchange with a cross section of 65 pb, i.e. a factor four smaller than the pair produced top quarks. In addition, the background, mainly W + jets, is rather

Fig. 31 Generic Feynman diagrams for single top production

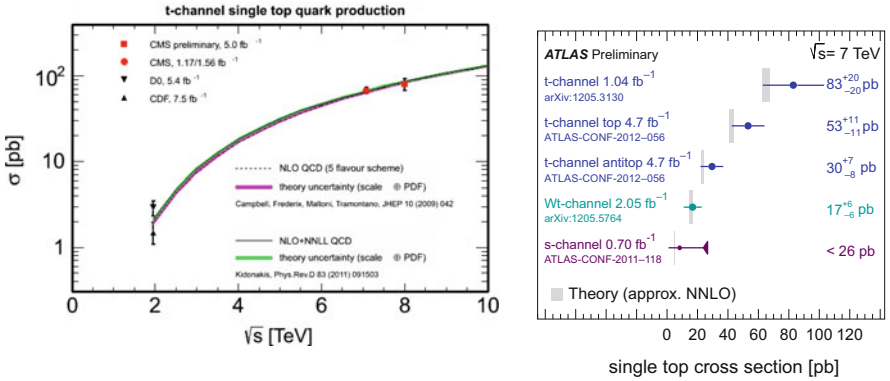

Fig. 32 Cross section for single top t-channel production at various centre-of-mass energies [80] (*left*) and LHC measurements for the different contributions to single top quark production [81]

high and tight cuts have to be applied to isolate a signal. Alternatively multi-variate analyses can be used.

Measurements have been performed both at the Tevatron and the LHC. The basic results are presented in Fig. 32. Both the energy dependence of the t – channel process [80] and the separation into different contributions agree with the Standard Model expectations [81].

The ratio of measured and expected cross sections can be used to obtain

$$| V_{tb} | = \sqrt{\frac{\sigma_{t-ch}}{\sigma^{th}_{t-ch}}} = 0.96 \pm 0.08 \pm 0.02 \tag{55}$$

where the second error is the experimental system one, the third the theory error. This can be interpreted as $V_{tb} > 0.81$ at 95 % confidence [80].

9 Conclusion

After 2 years of LHC data taking and 40 years of high precision challenges, the Standard Model is in excellent shape. All LHC measurements are in agreement with expectations, even though ATLAS and CMS extended the mass scales of Standard

Model tests by almost an order of magnitude. Instead, with the discovery of the Higgs boson with properties that agree with its expectations – at least within current experimental accuracies – the Standard Model seems to be complete and stronger than ever.

On the other hand, Standard Model studies at the LHC have just started. Many of the analyses are still based on just the 2010 data, while by now almost a factor of 1,000 more data have been collected. Many important studies will soon be made with higher statistics and reduced systematic uncertainties. Such studies are essential to use Standard Model processes as a tool to reach out for new effects and particles. Indeed, new effects may show up first in deviations from Standard Model expectations.

Once the LHC energy is ramped up to 13–14 TeV and produces yet higher luminosities, the range of Standard Model studies will be even more extended. The main challenge then is to keep the systematic knowledge under control. Once this is achieved, even more will Standard Model analyses provide the key to advancing our understanding. They may very well lead us into even more exciting times than those we have had recently.

Acknowledgements I would like to thank the organisers of the Scottish Universites Summer School on Physics, especially Craig Buttar and Franz Muheim, for inviting me to this exciting school. It was a real pleasure to have the many discussions with so many high quality students. I very much enjoyed also for being introduced to Scottish history, landscape and spirits.

References

1. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012)
2. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **716**, 30 (2012)
3. G. Aad et al., ATLAS Collaboration, JINST **3**, S08003 (2008)
4. S. Chatrchyan et al., CMS Collaboration, JINST **3**, S08004 (2008)
5. R.L. Gluckstern, Nucl. Instrum. Methods **24**, 381 (1963)
6. S. Chatrchyan et al., CMS Collaboration, JINST **7**, P10002 (2012)
7. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-043 (2012)
8. J.M. Campbell, J.W. Huston, W.J. Stirling, Rept. Prog. Phys. **70**, 89 (2007)
9. J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012)
10. P. Mättig, Phys. Rept. **177**, 141 (1989); I.G. Knowles, G.D. Lafferty, J. Phys. G **23**, 731 (1997)
11. T. Sjostrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008); T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001)
12. G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, B.R. Webber, JHEP **0101**, 010 (2001)
13. M.H. Seymour, M. Marx, (2013), arXiv:1304.6677 [hep-ph]
14. G. Aad et al., ATLAS Collaboration, New J. Phys. **13**, 053033 (2011)
15. V. Khachatryan et al., CMS Collaboration, Phys. Rev. Lett. **105**, 022002 (2010); B. Abelev et al., ALICE Collaboration, JHEP **1207**, 116 (2012)
16. V.M. Abazov et al., D0 Collaboration, Phys. Rev. D **81**, 052012 (2010); G. Aad et al., ATLAS Collaboration, New J. Phys. **15**, 033038 (2013)
17. M.G. Albrow et al., TeV4LHC QCD Working Group Collaboration, hep-ph/0610012, 2006

18. R. Field, Ann. Rev. Nucl. Part. Sci. **62**, 453 (2012)
19. S. Chatrchyan et al., CMS Collaboration, JHEP **1109**, 109 (2011)
20. S. Chatrchyan et al., CMS Collaboration, Eur. Phys. J. C **72**, 2080 (2012)
21. F.D. Aaron et al., H1 and ZEUS Collaboration, JHEP **1001**, 109 (2010)
22. M. Cacciari, G.P. Salam, G. Soyez, JHEP **0804**, 063 (2008)
23. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-PFT-10-002, 2010
24. G. Aad et al., ATLAS Collaboration, Eur. Phys. J. C **73**, 2304 (2013)
25. S. Chatrchyan et al., CMS Collaboration, JINST **6**, P11002 (2011)
26. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. D **87**, 112002 (2013)
27. G. Aad et al., ATLAS Collaboration, JHEP **1405**, 059 (2014)
28. G. Aad et al., ATLAS Collaboration, JHEP **1301**, 029 (2013)
29. E. Eichten, K.D. Lane, M.E. Peskin, Phys. Rev. Lett. **50**, 811 (1983)
30. V. Khachatryan et al., CMS Collaboration, Phys. Rev. Lett. **106**, 122003 (2011)
31. V.M. Abazov et al., D0 Collaboration, Phys. Lett. B **718**, 56 (2012)
32. M.H. Seymour, Z. Phys. C **62**, 127 (1994)
33. G. Aad et al., ATLAS Collaboration, JHEP **1205**, 128 (2012)
34. T. Aaltonen et al., CDF Collaboration, Phys. Rev. D **85**, 091101 (2012)
35. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **720**, 32 (2013)
36. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-SMP-12-011, 2012
37. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. Lett. **109**, 111806 (2012)
38. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-EXO-12-061, 2012; S. Chatrchyan et al., CMS Collaboration, CMS-PAS-EXO-12-060, 2012
39. F.A. Berends, W.T. Giele, H. Kuijf, R. Kleiss, W.J. Stirling, Phys. Lett. B **224**, 237 (1989)
40. S. Chatrchyan et al., CMS Collaboration, JHEP **1201**, 010 (2012)
41. G. Aad et al., ATLAS Collaboration, JHEP **1307**, 032 (2013)
42. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **708**, 221 (2012)
43. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-SMP-12-003, 2012; S. Chatrchyan et al., CMS Collaboration, JHEP **1206**, 126 (2012); G. Aad et al., ATLAS Collaboration, Phys. Lett. B **706**, 295 (2012)
44. G. Aad et al., ATLAS Collaboration, JHEP **1306**, 084 (2013)
45. Z. Bern, G. Diana, L.J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, S. Hoeche, H. Ita et al., Phys. Rev. D **84**, 034008 (2011)
46. G. Aad et al., ATLAS Collaboration, Eur. Phys. J. C **72**, 2001 (2012)
47. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **718**, 752 (2013)
48. S. Schael et al., ALEPH and DELPHI and L3 and OPAL and LEP Electroweak Collaborations, Phys. Rept. **532**, 119 (2013)
49. V.M. Abazov et al., D0 Collaboration, Phys. Rev. Lett. **108**, 151804 (2012)
50. T. Aaltonen et al., CDF Collaboration, Phys. Rev. Lett. **108**, 151803 (2012)
51. The Tevatron Electroweak Working Group, CDF and D0 Collaborations, (2012), arXiv:1204.0042 [hep-ex]
52. H. Aihara, T. Barklow, U. Baur, J. Busenitz, S. Errede, T.A. Fuess, T. Han, D. London et al., hep-ph/9503425, 1995
53. G. Aad et al., ATLAS Collaboration, Phys. Rev. D **87**, 112001 (2013)
54. T. Aaltonen et al., CDF and D0 Collaborations, Phys. Rev. D **86**, 092003 (2012)
55. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-TOP-12-003, 2012; S. Chatrchyan et al., CMS Collaboration, CMS-PAS-TOP-12-007, 2012
56. M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, M. Wiedermann, Comput. Phys. Commun. **182**, 1034 (2011)
57. M. Cacciari, M. Czakon, M. Mangano, A. Mitov, P. Nason, Phys. Lett. B **710**, 612 (2012); N. Kidonakis, Phys. Rev. D **82**, 114030 (2010); Phys. Rev. D **84**, 011504 (2011), arXiv:1205.3453
58. M. Czakon, P. Fiedler, A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013)
59. CMS Collaboration, http://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryPlots

60. S. Chatrchyan et al., CMS Collaboration, Eur. Phys. J. C **73**, 2339 (2013)
61. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 142 (2012); G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2011-153; S. Chatrchyan et al., CMS Collaboration, CMS-PAS-TOP-12-014, 2012
62. G. Aad et al., ATLAS Collaboration, Eur. Phys. J. C **72**, 2083 (2012)
63. G. Aad et al., ATLAS Collaboration, JHEP **1209**, 041 (2012)
64. R. Frederix, F. Maltoni, JHEP **0901**, 047 (2009)
65. G. Aad et al., ATLAS Collaboration, Phys. Rev. D **88**, 012004 (2013)
66. S. Chatrchyan et al., CMS Collaboration, JHEP **1212**, 105 (2012)
67. T. Aaltonen et al., CDF and D0 Collaborations, Phys. Rev. D **86**, 092003 (2012)
68. P.Z. Skands, D. Wicke, Eur. Phys. J. C **52**, 133 (2007)
69. S. Frixione, B.R. Webber, JHEP **0206**, 029 (2002)
70. M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, JHEP **0307**, 001 (2003)
71. V.M. Abazov et al., D0 Collaborations, Phys. Lett. B **703**, 403 (2011)
72. S. Moch, P. Uwer, Phys. Rev. D **78**, 034003 (2008); U. Langenfeld, S. Moch, P. Uwer, Phys. Rev. D **80**, 054009 (2009)
73. V.M. Abazov et al., D0 Collaboration, Phys. Lett. B **703**, 422 (2011)
74. S. Alioli, P. Fernandez, J. Fuster, A. Irles, S.-O. Moch, P. Uwer, M. Vos, Eur. Phys. J. C **73**, 2438 (2013)
75. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2011-054, 2011; S. Chatrchyan et al., CMS Collaboration, CMS-PAS-TOP-11-008, 2011
76. C. Quigg, Rept. Prog. Phys. **70**, 1019 (2007)
77. G. Degrassi, S. Di Vita, J. Elias-Miro, J.R. Espinosa, G.F. Giudice, G. Isidori, A. Strumia, JHEP **1208**, 098 (2012)
78. G. Aad et al., ATLAS Collaboration, JHEP **1206**, 088 (2012)
79. G. Aad et al., ATLAS Collaboration, Phys. Rev. Lett. **108**, 212001 (2012)
80. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-TOP-12-011, 2012
81. ATLAS Collaboration, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSumm-aryPlots/TOP/

Higgs Boson Searches

William Murray

Abstract This report reviews the discovery of a new particle in the search for the Standard Model Higgs boson by the ATLAS and CMS collaborations using 10 fb^{-1} of LHC data. Both experiments have observed a new boson with convincing evidence for the described. Some details of the $ZZ \to \ell\ell\ell\ell$ mode are entered into as an example while other modes are outlined more briefly. The implications and outlook are also summarised, along with some comments on more complex Higgs boson sectors.

1 Introduction

The first big discovery of the LHC programme was announced on the fourth of July 2012: the observation by ATLAS [1] and CMS [2] of a new particle found in the Higgs boson [3–5] search [6, 7]. At first sight this does appear to be compatible with the expectations for the Standard Model Higgs boson, and indeed the CERN Director General went so far as to refer to it as 'A Higgs boson'. Section 3 has a brief summary of the statistical methods, and in Sect. 4 the LEP and Tevatron contributions are summarised. Sections 5 and 6 discuss the evidence for the new particle, and Sect. 7 speculates a little on the next developments.

This note presents the public results when the lectures were given; using up to almost 5 fb^{-1} of 2011 7 TeV LHC data and in many cases the first quarter of the 2012 8 TeV data, a little over 5 fb^{-1}. In the lectures I spent some time discussing luminosity and emittance. This material is somewhat tangential to my main subject, and is well covered elsewhere [8], so it is not reproduced here. As I personally work on ATLAS and know the ATLAS analyses best, most of my examples will be drawn from there.

In this document, 'lepton' (ℓ) should normally be interpreted as referring to electrons or muons and their antiparticles. Limits are all quoted at 95 % CL.

W. Murray (✉)
RAL, Harwell Science and Innovation Campus, Harwell, OX11 0QX, UK
e-mail: bill.murray@stfc.ac.uk

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_5

2 Reminder of the Brout-Englert-Higgs Model

The Brout-Englert-Higgs mechanism, the introduction of a complex doublet field with a quartic self-coupling and negative quadratic term, has been described many times; a good example for experimentalists is Ref. [9]. I only comment on some of the historical development here.

The realisation that a spontaneously broken symmetry showed a route to evade Goldstone's theorem [10] and produce massive vector bosons was developed by a well-known series of papers [3–5] in 1964. One piece of the puzzle that was uniquely the insight of Peter Higgs [4] was the realisation that a massive boson was a consequence. The development was not widely celebrated until the formulation of the Standard Model in 1967 [11–13] and the proof that is renormalizable [14] in 1971. The audacity of this model may be missed today: of the 12 quarks and leptons a mere handful had been experimentally confirmed (e,μ,ν with hints for the u and d quarks), and the only force for which the vector boson was known was electromagnetism. However, with the discovery of the J/ψ and τ in 1973 the theory was widely accepted and continues to cover all the relevant experimental data today, with the honourable exception of neutrino oscillations.

The Higgs boson was initially viewed as an important but inaccessible part of the theory [15], but with the advent of LEP, the Tevatron and now the LHC it gradually assumed the role of 'most wanted' suspect in the Standard Model list. In the meantime, the discovery of the vector bosons and matter fermions, culminating with the tau neutrino, ν_τ, in 2001 [16], and the work of the SLC, LEP and the Tevatron in over-constraining the parameters, especially in the electroweak sector, led to a growing confidence in the Standard Model, at least as a low-energy effective theory.

Crucial to allowing the search to be undertaken seriously was the theoretical calculations of the Higgs production and decay properties [17, 18]. The most important production processes for hadron colliders are shown in Fig. 1, and although this is by no means a complete list (production associated with t or b quarks for example is omitted) it is in fact true that the discovery relies essentially on the first process only: ggF or gluon fusion.

The Higgs branching ratios and the cross-sections of experimentally useful decay modes are given in Fig. 2 [18]. It can be seen that one inverse femtobarn of LHC data contains, before experimental selections, thousands of Higgs bosons in the most prolific modes studied, and fewer than ten in the four-lepton decay channel. The problem, of course, is the enormous background rate.

The ggF process is a loop diagram that contributes about 90 % of the total LHC Higgs boson cross-section. Owing to overwhelming background from multijet production, most Higgs boson decay modes are probably not distinguishable from background in this channel, but those with highly distinctive signatures, the WW, ZZ and $\gamma\gamma$ channels, derive most of their sensitivity from it. It is calculated with 15 % precision at next-to-next-to-leading order (NNLO), and the difference in predicted cross-section at this order, compared with the leading order estimate,

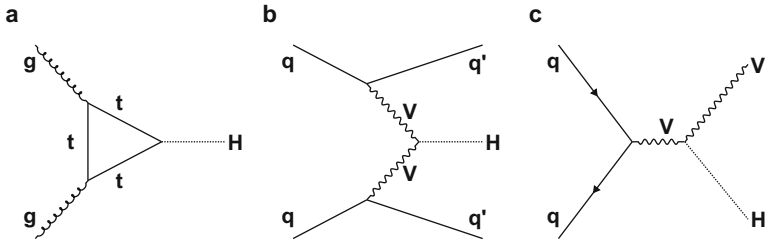

Fig. 1 The main Higgs boson production mechanisms at hadron colliders: (**a**) gluon fusion, denoted ggF, (**b**) vector boson fusion or qqH, (**c**) vector boson associated production or VH. This last diagram, if the quarks are replaced with leptons, also demonstrates the main production mechanism at LEP

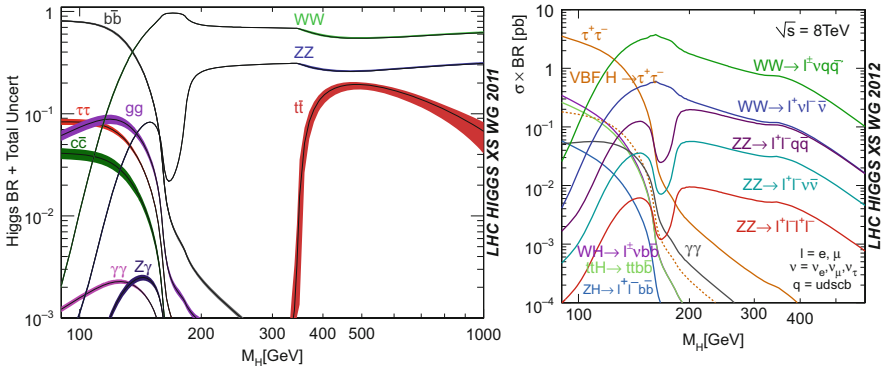

Fig. 2 *Left*: The main Higgs boson decay branching ratios as a function of mass [18]. *Right*: The cross-sections at the LHC running at 8 TeV, in various interesting decay modes [18]

is about a factor 2 increase in the rate. This big factor is at least in part because the original process involves a loop of coloured objects, so there are many possible gluonic corrections. Also, the momentum in the loop enhances high-p_T jets.

The second process, vector boson fusion (VBF), is of great experimental interest because the rate is not tremendously suppressed compared with ggF, but the two partons, recoiling from the vector boson emission present a distinct feature of relatively forward (backward) jets. This can be used to tag the process and enhance the signal to background. In particular there is no colour flow between the proton remnants and the signal to background ratio can be enhanced by vetoing on hadronic activity in regions of the detector between the tagging jets. A direct measurement of a control process such as Z production in VBF is greatly to be desired, but is not easy because of the large rate of Z production is association with jets from other sources. The $H \rightarrow \tau\tau$ searches particularly try to exploit this mode. The triggers used so far require leptons or photons from the Higgs candidate decay, but in principle it is possible to use triggers based on exploiting the jet topology.

Vector boson associated production is third in order of rate, a factor 20 or more below gluon fusion, at the LHC, but second at the Tevatron for low mass Higgs bosons. This reflects the kinematic suppression of the high-mass full VBF state at the lower Tevatron energy. The leptonic vector boson decay modes can be used to trigger the events, and the decay $Z \rightarrow \nu\nu$ can also be used if the Z has high enough p_T for the missing energy to be distinct. The VH mode has been exploited to good effect in the Tevatron $H \rightarrow b\bar{b}$ searches and the LHC experiments are also employing it, at least in part because there is a larger fraction of high p_T events in signal than in background, although this suffers from a very low rate.

The fourth production mode used so far is ttH, where the large top quark mass enhances the fraction of Higgs events and its distinctive decay signatures provide an effective tag. However, the event rates are low at 7 or 8 TeV and the jet radiation and detector complexities associated with double top decay make a significant observation in this mode very challenging. However, measuring the (fermionic) ttH production process directly will give a very useful comparison with the gluon-fusion process, which is a expected to be dominated by a top-quark loop. The production in association with a single top quark is strongly suppressed in the Standard Model, but would be enhanced if the sign of the coupling to fermions were inverted, and an analysis of this mode should be able to measure that sign.

The simplest version of the Higgs mechanism is employed in the SM, but it can be extended in many ways. One of the interesting extensions is the addition of a second Higgs doublet, and in particular the so-called 'type II' doublet [19] required by supersymmetry. Within the SUSY framework there are five physical Higgs scalars, two charged and three neutral, whose properties are completely defined at tree level by two parameters, often taken to be m_A and $\tan\beta$ (the ratio of the vacuum expectation values of the two doublets). The three neutral bosons are the lighter and heavier scalars, respectively h and H, and the pseudo-scalar A.

3 Statistical Aside

Much has been written on the statistical treatment used for the Higgs search [20–23]. As it happened the journey from evidence to discovery was faster than many had expected and therefore the statistical issues relating to the establishment or exclusion of marginal signals were not so crucial. Nevertheless a quick review may be useful.

The analyses (whether at LEP, the Tevatron or LHC) are done by considering one specific putative Higgs boson mass at a time. The CLs [22] method is generally adopted for expressing limits. CLs is the ratio of the probability of getting such a low signal in the data in the case of signal plus background, to the same probability for background only. The LHC experiments agreed to use CLs for their primary results, and the Tevatron experiments use it along side Bayesian limits with a flat prior. These approaches are identical in the case of a one bin (counting) experiment without systematic errors, and in practice give very similar results for the cases studied.

The evidence for a signal is assessed using a p-value. This is the fraction of times that the likelihood ratio calculated on the data events (or a value of the likelihood ration more extreme), would be expected in the absence of a signal. This is, by construction, a *local* p-value: that is to say it is the probability of observing such a result at a specific Higgs boson mass. The analyses of course consider a wide range of masses and therefore it is more likely that an extreme result is observed somewhere in the search range than at any particular mass. This can be allowed for when a total probability is quoted, but only by making some assumption about what mass range is interesting to consider. For example, are only masses where the SM Higgs boson is not excluded interesting? Or perhaps only masses where the experiment should be sensitive. As such questions do not have universally agreed answers the experiments quote the local p-values as the primary result.

The situation is further complicated by systematic errors. In the LHC analyses [20], these are treated by building a mathematical model where the systematic uncertainties are transformed into extra parameters of the model, known as nuisance parameters. These are then fitted along with the parameter of interest, the Higgs strength relative to the SM, μ, at the mass currently being considered. They are constrained by either a data control region (e.g. in the case of a background rate) or a constraint term (e.g. for a PDF uncertainty). When toy simulations are performed to test limits and p-values, the central values of the nuisance parameters are fixed, but the constraints are randomised to simulate an experiment measuring control regions multiple times.

Because these nuisance parameters are fitted in the analysis it is very important to construct the model that describes them with great care. If there is only one nuisance parameter assigned to electron energy scale, for example, then the presence of a high-statistics control region, such as a Z peak, in one of the distributions will fit this parameter rather precisely. This reduced scale error will then be applied to all other electrons in the analysis. In truth the jet energy scale depends upon the lepton η and p_T etc., and so a measurement with one sample of electrons cannot so easily be applied to another. While it may be reasonable to *describe* the energy scale error with just one number that does not mean that it can be fitted with one measurement. The LHC experiments have been gradually expanding their models to allow for this, and the ATLAS combined SM Higgs results at 125 GeV have over 1,000 nuisance parameters. Most of these in fact reflect limited simulation statistics in channels where the shape of a background is taken directly from simulation.

4 Prior Knowledge

Most of the information on the Higgs boson prior to LHC came from the LEP and Tevatron experiments, and some highlights of those results are presented here.

4.1 The Role of LEP

The LEP experiments were by no means the first to search for the Higgs boson, but they were in the position of being the first to study large numbers of Z bosons, to which the Higgs boson has a particular affinity, and thus give comprehensive results [24, 25]. The term 'Higgs-strahlung' was coined to describe the process where a Z radiates a Higgs boson, with either the initial state or final state Z being off shell, see Fig. 1c. The significant rate for this process, combined with the clean environment of e^+e^- collisions, allowed a comprehensive study up to the kinematic limit for the Higgs-strahlung process.

The first phase of LEP, LEP 1, collided electrons at the Z pole, where the large resonant Z cross-section, 30 nb, meant that 13M hadronic Z bosons were recorded by the four experiments together despite only a moderate instantaneous luminosity of $3.4\times10^{30}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The Higgs boson decay branching ratios are mass dependent, and successively higher masses were excluded through consideration of a stable Higgs boson or decay to $\gamma\gamma$, ee, $\mu\mu$, $\pi^+\pi^-$, $\tau\tau$ and $b\bar{b}$. These were studied in conjunction with clean decays of the Z to leptons (ee or $\mu\mu$) or neutrinos. No sign of a signal was observed, and it was concluded that the Higgs boson could not have a mass less than 65 GeV.

A slightly less sensitive approach, but much more general, is to concentrate on the decay of the recoiling Z. In e^+e^- collisions, unlike pp, the initial 4-momentum is known, allowing the measurement of the leptons from the Z decay to be used to calculate the mass of the recoiling object without reconstructing it directly [26]. This then meant that the Higgs boson could be excluded irrespective of its decay modes. This search mode may be very useful in measuring the total Higgs boson width at a future e^+e^- facility.

From 1995 to 2000 the LEP centre-of-mass energy was progressively raised from 91 GeV to over 200 GeV, and the most important production mode involved the final Z boson being on mass shell. This gave access to a Higgs boson with a mass up to approximately $\sqrt{s}-m_Z-2$, with a rather limited dependence upon luminosity, and the Higgs search thus motivated the highest possible beam energy. In the case of a circular electron collider the limit on the beam energy comes from bremsstrahlung losses as electrons are bent in the arcs, which rises as the fourth power of the relativistic γ factor. In LEP 2 some 4 % of the beam energy was radiated per turn.

The mass of the Z plays another special role in Higgs boson phenomenology: the lightest MSSM Higgs boson is constrained, at tree level, to be less massive than the Z boson, and when LEP was being considered this was known. Thus a 200 GeV collider was expected to either find the Higgs boson or disprove supersymmetry. However, as the top quark mass was discovered to be much higher than many people expected, and the loop corrections to the mass of the lightest supersymmetric Higgs boson were established, the limit was found to be about 135 GeV. John Ellis describes this calculation [27] as the most expensive of his life. Raising LEP's beam energy to access such masses would have involved a huge investment in super-conducting RF cavities to accelerate the beams and the physics case was

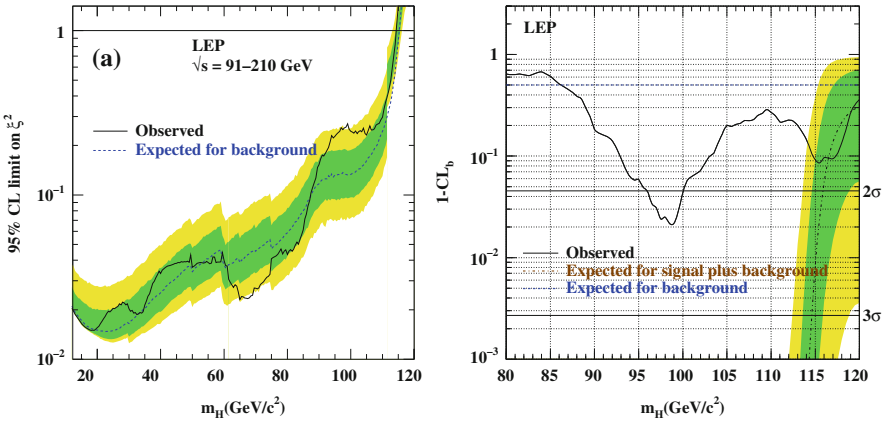

Fig. 3 *Left*: the limit on the Higgs boson production rate from LEP [24]. The y axis, ξ^2, is essentially the production rate in units of the SM strength. *Right*: the compatibility of the LEP data with the background as a function of m_H [24]. 1-CL_b is essentially the p-value

not considered overwhelming. Production of RF cavities was discontinued in 1996. The ultimate beam energy achieved was 208 GeV and LEP did not discover the Higgs boson. It is interesting to speculate how HEP might be different if that had happened 10 years ago, but the lesson we should draw now is to make the best use of each facility we have.

Given the beam energy, LEP should have set a limit of 115 GeV, but in fact 114.4 GeV was set [24] as shown in Fig. 3, owing to a small excess of Higgs boson-like candidates. This had a probability of occurring from background fluctuations, or p-value, of 0.09, as can be seen on the right hand side of the same figure. Such an excess should not be considered very unusual, but I was certainly one of the people disappointed that the LEP programme was not extended to see whether the Higgs boson had been found. Much has been said in subsequent years about the stronger excess at 98 GeV. However, while it could be argued that there was no relevant look-elsewhere effect for the Standard Model Higgs boson, which if it were lighter would have been discovered, and there was no sensitivity to higher masses, for this 98 GeV excess one really should consider a high-resolution search across a 100 GeV region, and in that context a 2 % probability is not in the least surprising.

4.2 The Part Played by the Tevatron

The closure of LEP at the end of 2000 coincided with the restart of the Tevatron, which had collected 0.11 fb^{-1} in what was called Run 1, between 1987 and 1995. The centre-of-mass-energy was raised from 1.8 to 1.96 TeV, but more importantly Run 2 delivered approximately 10 fb^{-1}, 100 times more data than Run 1. The bunch

crossing time of 175 ns meant that pileup was several times worse than at LHC for an equivalent luminosity, and many groups joined the Tevatron experiments not only to make great measurements and perhaps big discoveries, but also to learn how to do collider physics in a high pileup environment, in preparation for the LHC.

The Higgs search at the Tevatron is sensitive for Higgs boson masses between 100 and 200 GeV. For masses below 130 GeV sensitivity is largely through VH production, where V represents either a W or a Z boson, and the Higgs decays to $b\bar{b}$. For higher masses the WW decay mode is primarily employed.

The Tevatron integrated luminosity took longer to achieve than had been hoped – but the LHC was delayed significantly in the meantime. The result was that a huge swathe of physics was done at the Tevatron, with discoveries like B_s oscillations [28] and incredible precision achieved on the W [29, 30] and top masses [31]. The slow start, coupled with the fact that the collider had run at essentially the same energy in Run 1, meant that there was a long period when no-one expected to discover new physics and instead effort was invested in understanding the detectors and learning how to get the best out of them.

In the Higgs search in the di-boson channels, essentially WW at the Tevatron, the signal is generated largely through the loop process of gluon fusion, Fig. 1, while the backgrounds can exploit tree level quark annihilation diagrams. The VH diagrams are essentially driven by quark annihilation for both signal and background. Had the Tevatron been a pp collider rather than $p\bar{p}$, the quark annihilation diagrams would have been suppressed. This would have reduced the signal in the VH modes, but the gluon loop signal processes would have benefitted from reduced di-boson background.

When datasets of many fb^{-1} became available, the Tevatron experiments, following normal practice, took to updating their results for summer and winter conferences. These updates were often partial but present an interesting historical perspective. One trend which is very marked is that with (almost) every update the sensitivity improves more than would be expected by projecting the previous analysis to a larger dataset. In other words, familiarity with the conditions and the detector, and the continual flow of good ideas, seems to overcome the systematic errors and it seems reasonable to assume a similar feature may occur in LHC searches; as indeed it appears to be doing. However, it must be emphasised that this improvement comes from a very large investment of effort in lots of detailed work.

The current results [32], which are close to final, are shown in Fig. 4. Between 120 and 130 GeV, the p-value is around 2×10^{-3}, which is indicative of a signal strength about double that expected for the SM Higgs boson. A separate publication [33], considers the bb final state separately and finds a p-value of 7×10^{-4} at 135 GeV, or 2×10^{-3} at 125 GeV.

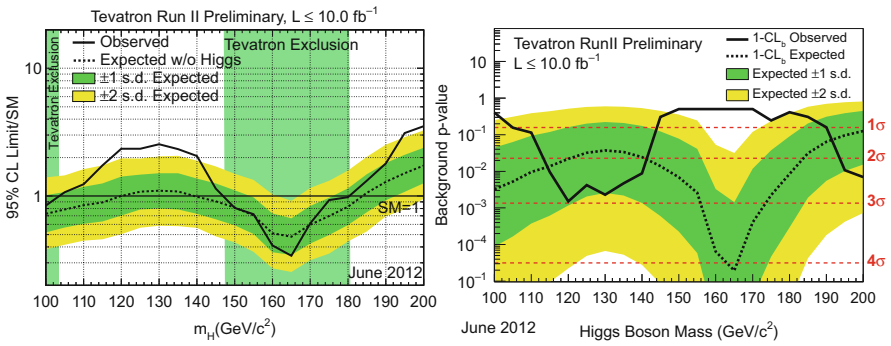

Fig. 4 *Left*: The upper limit on the Higgs boson cross-section, in units of the SM signal strength, from the combination of Tevatron searches [32]. If the limit is below 1 then the Standard Model Higgs boson is excluded at that mass at at least 95 % CL. *Right*: The p-value for the Tevatron search [32]. Around 120–130 GeV the combined probability approaches 1 in 1,000

5 MSSM Higgs Bosons

In the MSSM, the search for the lightest scalar, h, is in most scenarios closely related to the SM Higgs boson search described later. This section sketches the results of the search for the other MSSM Higgs bosons, the neutral H and A and charged H^+.

5.1 Neutral MSSM Higgs Boson Searches

The heavy MSSM Higgs bosons, A and H, do not couple strongly to the W and Z bosons, but have a coupling to the down-type fermions proportional to $\tan\beta$. Searches at the LHC [34, 35] have so far focused on the decay mode H/A to $\tau\tau$, which has a sensitivity growing with $\tan\beta$. The production comes either through gluon fusion, dominant for low $\tan\beta$, or from associated production with one or more b quarks, which grows proportional to $\tan\beta^2$. The results from the ATLAS and CMS collaborations are shown in Fig. 5, using around 5 fb^{-1} of 7 TeV data and a detailed analysis of the production mode. That is to say, three production modes were considered: gluon fusion, b-quark associated, and vector boson fusion or VBF. The second is especially appropriate for high $\tan\beta$ MSSM Higgs bosons.

The τ pairs were studied in $e\mu$, $\mu\mu$, $e\tau_{had}$ and $\mu\tau_{had}$ decay modes, with no evidence seen for Higgs boson production. A very large region is excluded, especially for large $\tan\beta$. However, the region with m_A above about 250 GeV and moderate $\tan\beta$ such at 5 is known as the 'LHC wedge' and will be difficult to explore with searches of the style used here.

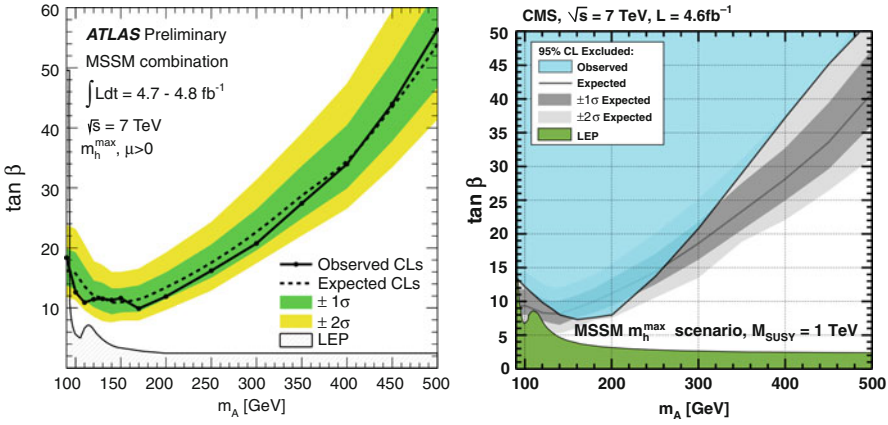

Fig. 5 Limits on the $m_A - \tan\beta$ plane coming from the H/A to $\tau\tau$ search in ATLAS [34] (*left*) and CMS [35] (*right*)

5.2 *Charged MSSM Higgs Boson Search*

The coupling between the charged Higgs and the top quark is strong, and most charged Higgs boson searches are in association with a top quark. The phenomenology depends crucially upon the relative masses, and hence which decays into which. If the charged Higgs weighs less than the top quark it can be produced with a large rate in top decay. If not, then the production cross-sections are not yet accessible at the LHC, although they may become so with the full Run 1 data. The mass in the MSSM is similar to the heavy neutral Higgs bosons.

Searches for charged Higgs bosons lighter than the top quark have been reported by both LHC collaborations. ATLAS produced limits [36] on the hadronic $c\bar{s}$ decay mode, but for most MSSM parameter choices the dominant decay of a light charged Higgs boson is $H^+ \to \tau\nu$, for which searches are reported [37, 38] by both CMS and ATLAS.

Both analyses consider multiple final states according to the decay modes of the taus from the charged Higgs decay and the W boson from the other top quark. The case where both the W and the tau decay to hadrons has the largest branching fraction and provides the tightest constraints, despite the background from multijet production. The ATLAS search has used a revised transverse mass to improve sensitivity.

The combined limits extracted from the channels are shown in Fig. 6 in terms of the fraction of top quarks decaying to charged Higgs, assuming that these all subsequently decay to $\tau\nu$. This fraction is found to be below 4 % or less for the masses tested. The implications are also shown in the $m_A - \tan\beta$ plane. The latter introduces more model dependence, and because of a minimum in the cross-section for $\tan\beta$ around 8 shows two separate excluded areas. Comparison of Figs. 5 and 6

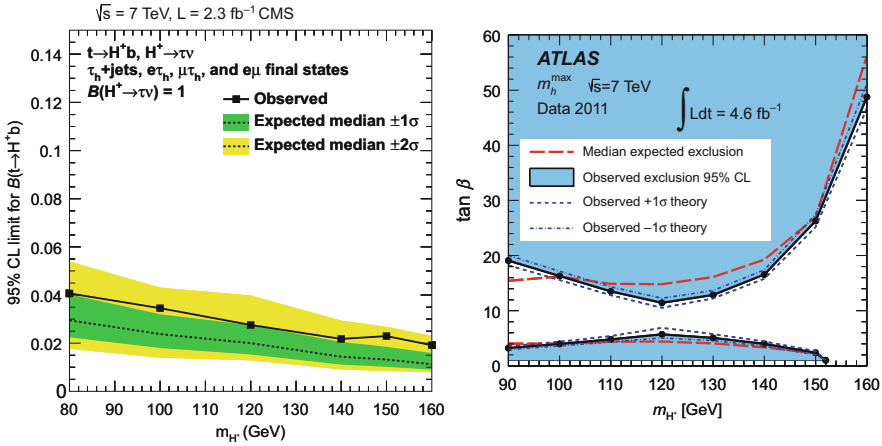

Fig. 6 *Left*: Limits on charged Higgs boson production from CMS [37], expressed in terms of the top branching ratio to charged Higgs boson. *Right*: The ATLAS search results [38], interpreted in the $m_A - \tan\beta$ plane

reveals that, at least in the m_h^{max} scenario, the charged Higgs searches have little sensitivity beyond the $H \to \tau\tau$. However, the charged Higgs boson search has the promise of an unambiguous indicator of physics beyond the Standard Model, and will in future exclude a region of m_{H+} for all $\tan\beta$. Furthermore, we should not limit ourselves to only looking for what today appears to be the best motivated theory.

6 SM Higgs Boson

A key element in the LHC Run 1 strategy was the conclusion of the search for the SM Higgs boson. The schedule was built around delivering enough luminosity to convincingly exclude or discover this particle. This was of course rewarded by the appearance of a new boson, but it is important to first consider the context of that discovery: the exclusion of essentially all other reasonable masses.

The only free parameter of the SM Higgs boson was its mass; after that everything is predicted. For Higgs boson masses above about 130 GeV the sensitivity is dominated by the bosonic decay modes, WW and ZZ. These have so far been searched for in channels where at least one vector boson decays to leptons: WW $\to \ell\nu\ell\nu$ [39, 40], WW $\to \ell\nu$qq [41, 42], ZZ $\to \ell\ell\ell\ell$ [43, 44], ZZ $\to \ell\ell\nu\nu$ [45, 46] and ZZ $\to \ell\ell$qq [47, 48]. For lower masses, down to the LEP limit at 114 GeV, the decays to $b\bar{b}$, $\tau\tau$ and especially $\gamma\gamma$ play a gradually increasing role. The results from the 2011 run [50, 51] are displayed in Fig. 7. Both experiments excluded a large range of Higgs boson masses, and really left only a small region, from 115 GeV

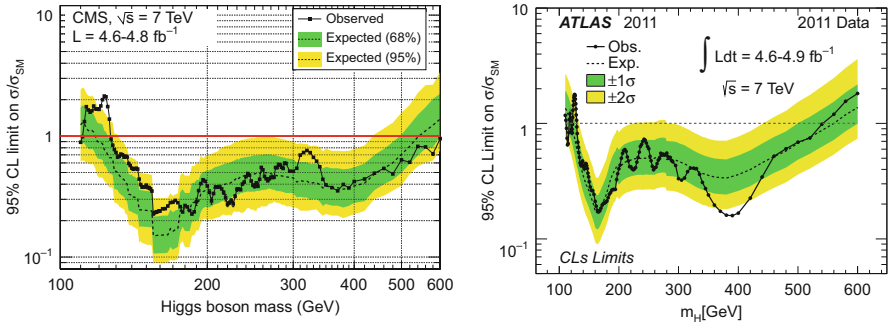

Fig. 7 Upper limits set by the CMS [50] (*left*) and ATLAS [51] (*right*) experiments on SM Higgs boson production, in units of the expected cross-section. Masses from 130 to 500 or 600 GeV are excluded, in line with expectations, but the failure to exclude around 125 GeV is already interesting

or so to 130 GeV in which the SM Higgs boson could be expected to be found. Technically a SM Higgs boson above 600 GeV was not excluded, but this is strongly disfavoured by the electroweak fits [52].

It is important to emphasize the background issues faced at LHC. At the time of discovery of order of 100,000 Higgs bosons had been produced in each of the large detectors, but these were almost overwhelmed by 8×10^{14} inelastic pp collisions. The sheer impossibility of recording more than a tiny fraction of the collisions makes the trigger systems of ATLAS and CMS critical. The discovery came in the end almost entirely from non-hadronic channels: electrons, muons, photons and to some extent neutrinos. These present distinct and triggerable signatures and finally allow the observations, including observing a mere dozen Higgs boson to four lepton candidates from that phenomenal background.

6.1 Higgs Boson Decay to ZZ

The decay to pairs of Z bosons has the potential for very powerful searches due to the attractive features of the subsequent Z boson decays: particle-antiparticle pairs allow cross-checks, and the presence of one or two Z boson peaks allows important cross-checks. ATLAS and CMS have both studied events where one Z decays to leptons and the other to leptons, neutrinos or quarks.

6.1.1 $H \rightarrow ZZ \rightarrow \ell\ell\ell\ell$

The cleanest channel used for the Higgs boson search at the LHC is the decay to pairs of Z bosons with their subsequent decay to electron or muon pairs [43, 44, 53, 54]. I shall describe this is in a little more detail than the others as

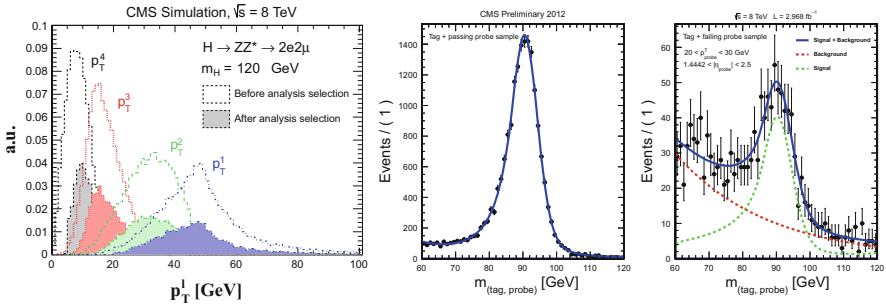

Fig. 8 *Left*: The expected p_T distributions of the four leptons produced by a decaying Higgs boson of mass 120 GeV, as seen in CMS simulation. *Centre* and *right*: The dilepton invariant mass distribution around the $Z \to ee$ peak seen when both leptons pass the identification criteria (*centre*) and when one of them fails (*right*)

an example of the sort of techniques used. However, the signal rate is very low and so the primary emphasis is on maximising the signal efficiency whilst preserving the low background. The multiple constraints of four clean leptons and one or two resonant Z bosons means that the selected sample is dominated by real di-boson production, with potentially a narrow Higgs boson signal in addition. To investigate a Higgs boson weighing less than twice the Z mass, one of the Z bosons must be off mass shell, and the mass window for the pair of leptons from it is therefore large. The mass resolution for a light Higgs boson is of order 2 GeV. There is no need to make requirements on the production process; an inclusive approach maximises the efficiency.

ATLAS restricts the analysis to electrons or muons while CMS also considers τ leptons, but despite more than half of leptonic ZZ decays featuring at least one τ pair the additional sensitivity they bring is small and I shall not discuss them further.

CMS imposes a minimum requirement on the mass of both di-lepton pairs of 12 GeV for events compatible with a 125 GeV Higgs boson, and ATLAS used 17.5 GeV. These exclude any contamination from upsilon or J/ψ decay, but are otherwise rather open. For the same reason it is important to use low p_T thresholds on the leptons, as can be seen in Fig. 8, where the lowest transverse momentum lepton of the four is typically below 10 GeV in p_T. Thresholds of 5–6 GeV in p_T are used for muons and 7 GeV for electrons. The efficiencies and fake rates of these leptons are measured in data by studying known resonance peaks, such as the Z to di-lepton. An example is given in Fig. 8 of the Z to di-electron peak reconstructed in CMS using events where an electron with p_T between 20 and 30 GeV, known as the probe, either passes or fails the identification criteria. Analyses using the J/ψ allows give access to lower momentum leptons.

The potential exposure to high rates of fake leptons is minimised with moderately tight lepton identification criteria. The most important backgrounds, like Z plus jets are suppressed by the requirement of four identified leptons. Secondary leptons from b quark decays are a particular issue, and these (as well as other backgrounds)

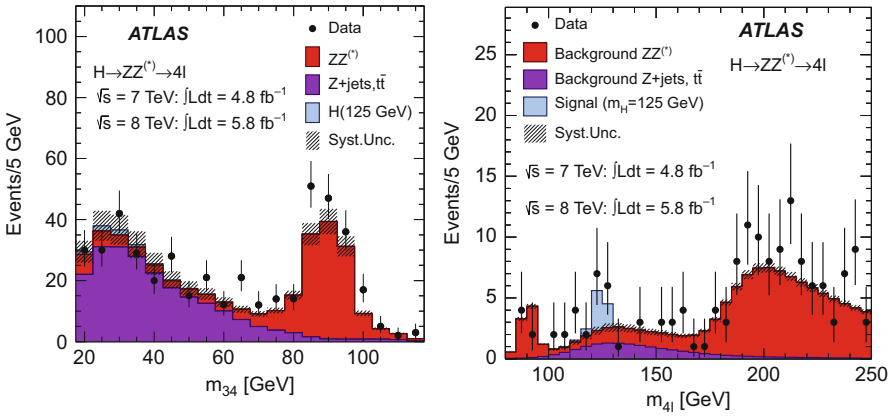

Fig. 9 Control distributions as recorded by ATLAS [6]. Events are selected as in the full analysis, but the lepton isolation and impact parameter requirements are not applied to the '34' lepton pair

are controlled using isolation and impact parameter requirements that target the associated hadrons and the visible flight distance of the b quark. The background rates too are assessed from the data.

Figure 9 shows the effect of releasing the isolation and impact parameter selections from the Z candidate with measured mass furthest from the nominal Z boson mass in the ATLAS analysis. The analysis is split by the flavour of the soft lepton pair, as electrons have a much larger contribution from fakes than muons do. The resonant peaks of single and di-boson production can be clearly seen and the off-resonance component is used to measure the background contributions. These are then used to estimate the background in the signal region, using the probability that a secondary lepton passes the full selections verified or measured in other control samples.

The background after applying all the selections is dominated by the irreducible non-resonant ZZ production, for which the shape is predicted by POWHEG [55] or MCFM [56]. This is mainly a $q\bar{q}$ production process, which is calculated at NLO, but in the low mass region has a 15 % contribution from gluon fusion processes which do not have the same precision. This region also has significant backgrounds from Z plus jets and $t\bar{t}$, which are constrained from the data as discussed. The observed mass spectra, and predicted backgrounds, can be seen in Fig. 10.

The data in each experiment are compatible with the sum of the known backgrounds, with the exception of a mass region near 125 GeV, where there is small, but statistically quite significant, excess of signal events. The peak of $Z \to \ell\ell\ell\ell$ can be clearly seen in the CMS distribution and is less efficiently selected by the ATLAS cuts. It is a very useful calibration point giving extra confidence in the efficiency and mass scale of the analyses. The details of the masses of the lepton pairs in the ATLAS candidate events with a four lepton mass between 120 and 130 GeV can be seen in Fig. 11. Three of the four events where both lepton pairs

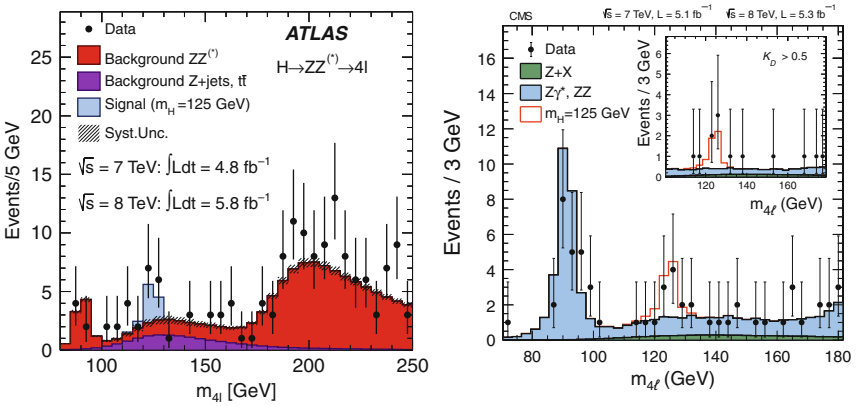

Fig. 10 The masses of the selected event candidates in the $H \rightarrow ZZ \rightarrow \ell\ell\ell\ell$ search of ATLAS [6] (*left*) and CMS [7] (*right*). The SM background is shown, including the single Z to four leptons peak at 91 GeV. The expected size and shape of a signal around 125 GeV is superimposed

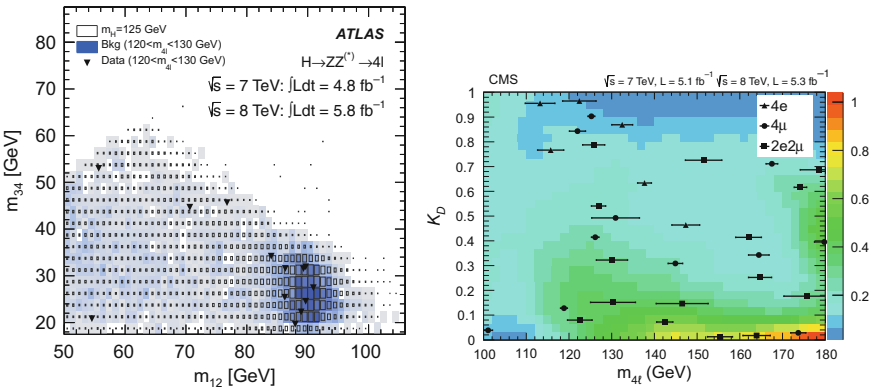

Fig. 11 Some kinematic details of the selected candidates. *Left* shows the distribution in ATLAS of the masses of the two Z boson candidates [6]. *Right* shows the distribution in CMS of the 'MELA' versus the four-lepton mass [7]

have masses below 80 GeV have rather large mass on the m_{34}, the lower lepton pair mass, which is more typical of signal than background.

CMS takes the analysis of the kinematic details of the events (not only the lepton-pair masses but also the decay angles) further, by constructing a likelihood ratio for Higgs signal versus four-lepton background as seen in Fig. 11; analysis of this yields a p-value for the background hypothesis below 10^{-3} for masses around 125 GeV, with the expected value in the presence of a signal being compatible but somewhat more extreme. ATLAS just fits the mass spectrum divided up by lepton species. This is a less sensitive search, but it happens to observe a p-value of about 3×10^{-4} in the same region, with an expectation perhaps 10 times less extreme.

In no other mass region is there evidence for a discrepancy passing the 1 % level in either experiment. This is not particularly surprising given the limits shown in Fig. 7. Thus the four lepton data gives a highly suggestive, but by itself inconclusive, hint for a new particle near 125 GeV.

6.1.2 $H \rightarrow ZZ \rightarrow \ell\ell qq$ or $ZZ \rightarrow \ell\ell\nu\nu$

The $\ell\ell\nu\nu$ [45, 46] and $\ell\ell$qq [47, 48] decay modes have larger branching ratios than the four lepton decay, but also allow good separation of a Higgs boson candidate from background. One Z boson is reconstructed through its decay to electrons or muons and the other is looked for either as a pair of jets or through missing transverse momentum. In each case the signal to background ratio is much improved if both Z bosons are on mass shell, and so these channels have mostly been used for Higgs boson masses above 200 GeV or more, although CMS have searched for the $llqq$ mode for masses down to 130 GeV [48]. They have not been used to test for a Higgs boson near 125 GeV. In the case of the $llqq$ decay a fair resolution on the Higgs boson candidate mass can be achieved, but the mode with two missing neutrinos does not have enough information to calculate the full invariant mass and a transverse mass is calculated instead, using the missing transverse momentum as a proxy for the Z momentum, and assigning the Z mass to it. As the expected Higgs boson width rises rapidly with mass, the fact that the experimental resolution is worse than in the four-lepton mode is of little importance at the highest masses tested, 600 GeV.

The total ZZ production cross-section is of order 7 pb at 8 TeV: much smaller than the inclusive Z plus backgrounds. Thus Z plus jets dominates the $llqq$ channel. In the case of $\ell\ell\nu\nu$ the missing energy resolution is such that the background at high Higgs boson masses is largely genuine ZZ production. In each case the selections made rely on simulation to model the signal acceptance, but the background estimation techniques are quite different.

The trigger and initial selection rely on selecting a lepton pair compatible with the Z boson mass. Then the analysis requires either two jets or significant missing transverse energy, depending upon the ZZ decay mode sought. For the di-jet case, the sensitivity is improved by tagging b-quarks as they are produced in a much larger fraction of ZZ decays than in the Z plus jets background. The experiments use b-tagged and untagged candidates separately.

The mass distributions in the (high-mass) b-tagged channel, which provides approximately half of the total sensitivity, can be seen in Fig. 12.

The ZZ $\rightarrow \ell\ell$qq and ZZ $\rightarrow \ell\ell\nu\nu$ searches do not reveal any striking excess, never departing from the two-sigma expected region. The $\ell\ell$qq analyses, using the 7 TeV data only, have a sensitivity which only reaches the SM signal strength in a small region of mass near 400 GeV. In contrast, the $\ell\ell\nu\nu$ result from ATLAS rules out a 200 GeV wide region based on the 7 TeV data, while that from CMS excludes a SM Higgs boson with mass between 273 and 600 GeV, using 10 fb^{-1} of 7 and 8 TeV data.

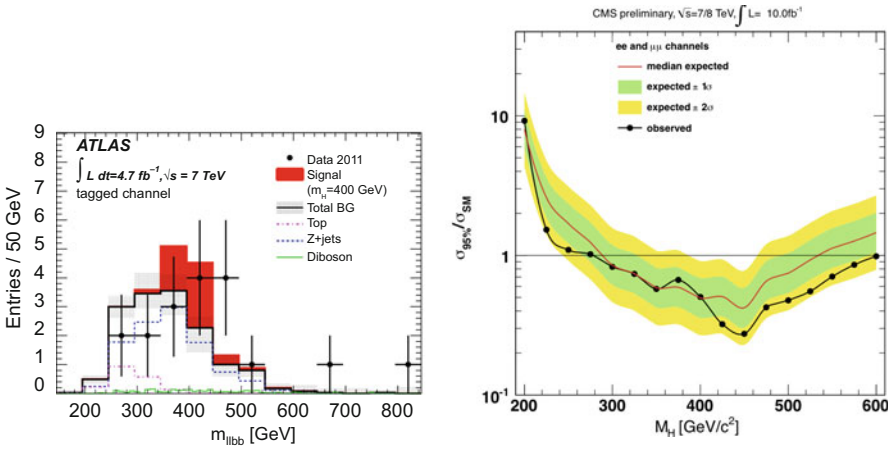

Fig. 12 The mass distribution measured by ATLAS in the $llbb$ mode [47] (*left*) and by CMS [46] in $\ell\ell\nu\nu$ decay (*right*)

6.2 *Higgs Boson Decay to $\gamma\gamma$*

The requirements of measuring the $\gamma\gamma$ decay process [49, 50, 57, 58] have driven the intended performance of the electromagnetic calorimeters of ATLAS and CMS. The CMS crystal calorimetry offers superior energy resolution while the ATLAS segmented calorimetry allows a simultaneous measurement of the photon angle. This means that the Higgs boson candidate decay vertex location in CMS requires the use of information from tracking system, while ATLAS can rely simply on the calorimetry; but that when the primary vertex can be cleanly identified, the CMS resolution is better. Maintaining the excellent resolution of a crystal calorimeter in high radiation environment has required continuous monitoring and recalibration. In practice the sensitivity of the two systems in this measurement is rather similar.

The analysis relies upon selecting two clean photons with transverse momenta above about 30 GeV, using shower shape and isolation criteria. In practice, asymmetric p_T cuts are used as these sculpt the mass distribution less near the turn on. The trigger for this channel relies on two photons identified online with looser and lower momentum selections, and is of order 99 % efficient in both experiments for events that would pass the offline selection criteria. As always, it is important to suppress large backgrounds, but both experiments reduce the dijet background to a few percent of the genuine diphoton rate. There is a more important contribution from photon plus jet events, which comprises 20 or 30 % of the sample. With this efficiency there is little to be gained by further reduction and the emphasis is on maintaining efficiency.

To improve sensitivity, the candidates are split into samples with different mass resolutions or signal to background ratios, depending for example upon the p_T of the Higgs boson candidate or the position of the detected photons in the calorimeter.

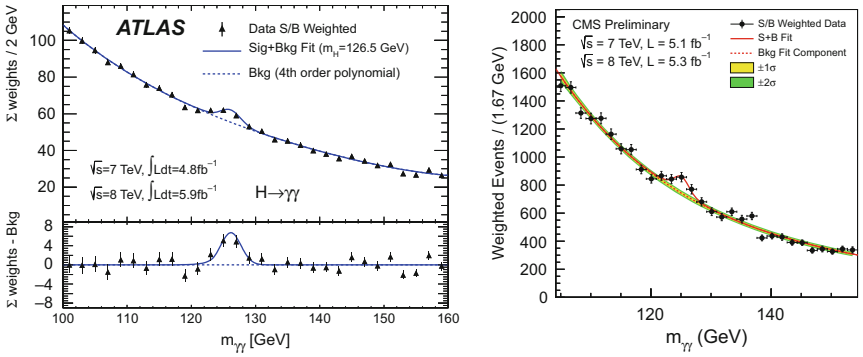

Fig. 13 The two-photon mass spectrum as measured by ATLAS [6] (*left*) and CMS (*right*)

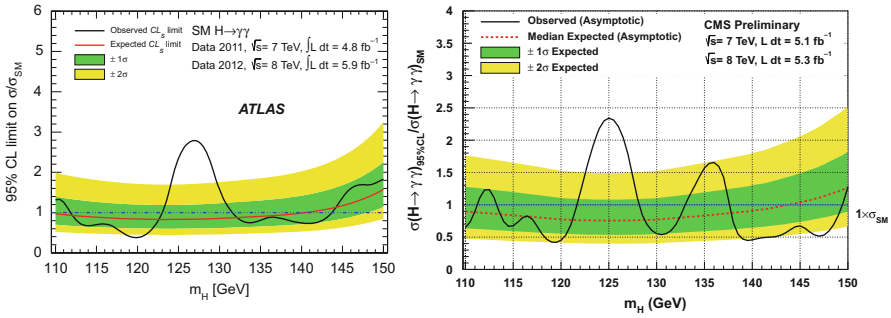

Fig. 14 The limits set on the SM Higgs boson cross-section in the $\gamma\gamma$ channel by ATLAS [6] (*left*) and CMS (*right*)

There are also one or two categories which require two jets with a large mass and separated in η which are designed to give sensitivity to the VBF production process. The complete mass spectra of both experiments are shown in Fig. 13, but with the separate samples weighted by the signal to background ratio and summed. This approximation presents all the candidates without having to show the many sub-distributions -20 in the case of ATLAS.

The signal is extracted by fitting the observed mass spectrum with a smooth function to describe the background plus a signal contribution. It is important to establish limits on the deviation between the true background shape and the simplified model; CMS requires these to be below a certain threshold while ATLAS parametrises the deviation as a 'spurious signal' which subtracts from any observed signal in the fit.

Each experiment has sufficient sensitivity to expect to exclude a SM Higgs boson from 110 to 140 GeV (or a little more in the case of CMS), as can be seen in Fig. 14. In each case, at 125 GeV the actual limit set is at over twice the SM signal strength, owing to the excesses clearly visible in Fig 13. The probability of these as background fluctuations, evaluated locally at 126 GeV, are of order 10^{-5} in each of

ATLAS and CMS, which is the strongest evidence for the new boson. The signal strength is 1 or 2 sigma more than expected, but we should not attach too much importance to deviations of 1 or 2 sigma.

It is important to calibrate the position of this peak accurately, both to know what the Higgs boson mass is and to establish the consistency or otherwise of the experiments and search channels. This is largely done using the related $Z \rightarrow ee$ mass spectrum. Electrons and photons interact rather similarly in the electromegnetic calorimeters, so this calibration applies to both. However, as electrons are charged they are more sensitive to the upstream material (especially when converted photons are treated separately, as is done here). There are detailed studies of this material, but there is a degree of reliance on Monte Carlo simulation to correct the scales. This is verified using $Z \rightarrow \ell\ell\gamma$ events to provide an absolute check on the photon energy scale, but there are not enough such photons at high p_T to provide a complete test.

6.3 Higgs Boson Decay to WW

The largest branching fraction for the SM Higgs boson is to W boson pairs, which approaches 100 % for masses around 165 GeV. This channel has been used to search for the Higgs boson in the case that one, or more usually both, W boson decays to lepton and neutrino. In each case a leptonic trigger can be used to select the events.

6.3.1 $H \rightarrow WW \rightarrow \ell\nu\ell\nu$

The first search mode to achieve SM sensitivity was the doubly leptonic WW decay, with maximum sensitivity near 165 GeV [39, 40]. The presence of two leptons and substantial missing energy allows for excellent multijet suppression, and the spin-zero nature of the Higgs boson aligns the spin of the Ws and hence, through maximal parity violation, the decay leptons tend to be emitted in similar directions [59], while the dominant WW background does not have this feature. The current analyses focus on the mass region near 125 GeV, where the low mass forces one of the W bosons to be off mass-shell. This weakens the kinematic correlation coming from the spin and with the low p_T lepton threshold required plus the lack of kinematic constraints due to the two missing neutrinos this becomes a very delicate channel to analyse.

Several different background sources contribute. The signal to background ratio achieved is of the order of 1–10, and very precise control of the background modelling is essential. Examples of background control or validation regions can be seen in Fig. 16.

The search is done by first reducing background sources without two vector bosons by the identification of two isolated opposite-charge leptons and missing energy. This selection is dominated by same-flavour lepton pairs from Drell-Yan

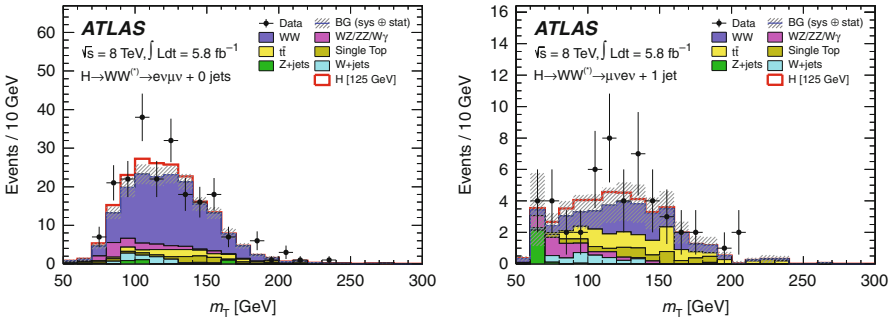

Fig. 15 Observed m_T distributions from the ATLAS $H \rightarrow WW \rightarrow lvlv$ search [6]. *Left* is for events with no extra jets, in the example of the muon being the subleading lepton, and *right* shows events with one extra jet when the electron is subleading

production with missing energy from experimental resolution. Thus the observation of one electron and one muon gives an easier signal extraction. The missing transverse energy requirements can reduce the Drell-Yan, but nevertheless most of the sensitivity comes from the opposite flavour channels, especially with the large pileup seen in 2012. ATLAS has only presented results using the opposite flavour channels so far in 2012. The events are divided into subcategories dependent upon the number of associated jets, and specific kinematic selections are made in each. The leptons are required to be close in azimuth to exploit the spin correlations mentioned above. As electrons suffer from higher fake rates than muons, particularly at low p_T, the events are split according to the flavour of the softer lepton.

In the 0 jets category the background is dominantly non-resonant WW, while the 1 jet events have a sizeable top contribution and in the exclusive two jet case, additional cuts on the rapidities of the jets are applied to enhance sensitivity to the vector boson fusion production process. As a final step the analyses fit the transverse mass of the leptons and the missing energy [60] as shown in Fig. 15.

The dominant background is non-resonant WW production, and this has to be estimated carefully. Control regions kinematically close to the signal but with higher m_{ll}, or with the leptons not close in azimuth, are used to measure the rate of WW production, and simulation is then used to extrapolate this rate into the signal region and model the distribution. This is shown for the ATLAS analysis in Fig. 16 (left). The modelling of the extrapolation from this control region into the signal region is a major systematic error. There are several other background contributions from di-bosons or top quarks which are important, but the largest uncertainty arises from $W +$ jets estimation, where the probability that a jet is mistaken for a lepton depends upon the initial parton type and there are therefore systematic errors related to the sample composition even when a control region is identified. One check of this estimation can be seen in the same-sign control region in Fig. 16 (right).

This channel has important sensitivity, with expected exclusions down to 125 GeV in ATLAS and 122 GeV in CMS. The observed exclusion starts rather

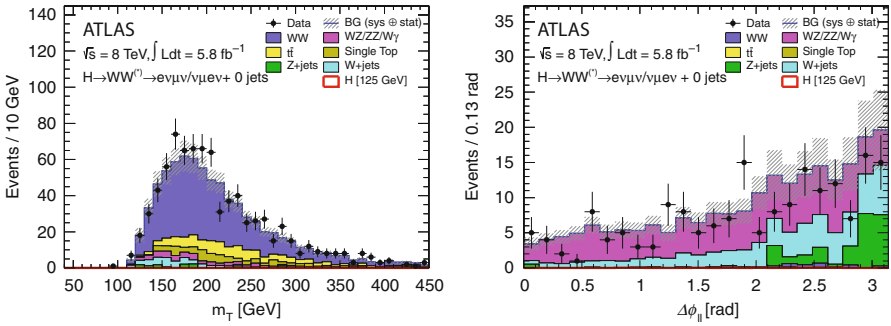

Fig. 16 Example background control distribution in the ATLAS WW $\rightarrow \ell\nu\ell\nu$ search [6]. *Left*: Non-resonant WW control region. *Right*: Events with two leptons of the same charge

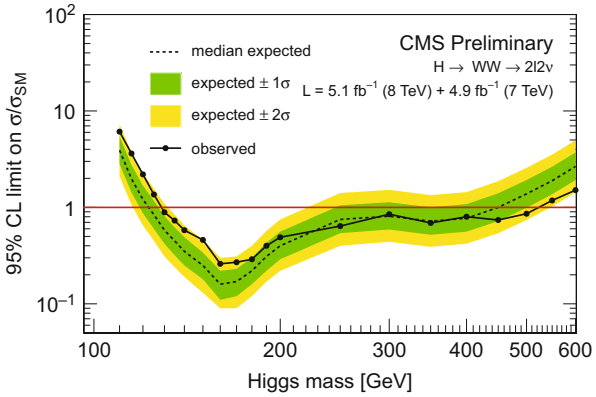

Fig. 17 Limits from CMS [40] on $H \rightarrow WW \rightarrow l\nu l\nu$ production as a function of m_H

higher, being 136 GeV in ATLAS and 129 GeV in CMS (shown in Fig. 17). The reason for the discrepancy between expected and observed at low mass is a significant excess of candidates in both experiments. The mass resolution is very poor, owing to the two missing neutrinos, and this excess manifests itself over a wide mass region. The systematic errors assigned to the background modelling have a major impact on the analysis, and improvements in sensitivity will need great care in the background studies.

6.3.2 $H \rightarrow$ WW $\rightarrow \ell\nu$qq

The largest Higgs branching ratio for high masses is WW $\rightarrow \ell\nu$qq [41, 42]. The three momentum of the neutrino can be estimated from the missing transverse energy and by imposing a W boson mass constraint on the $\ell\nu$ system, which allows the p_z to be found, up to an ambiguity in the roots of a quadratic equation. This allows the complete Higgs boson candidate mass to be calculated for each event.

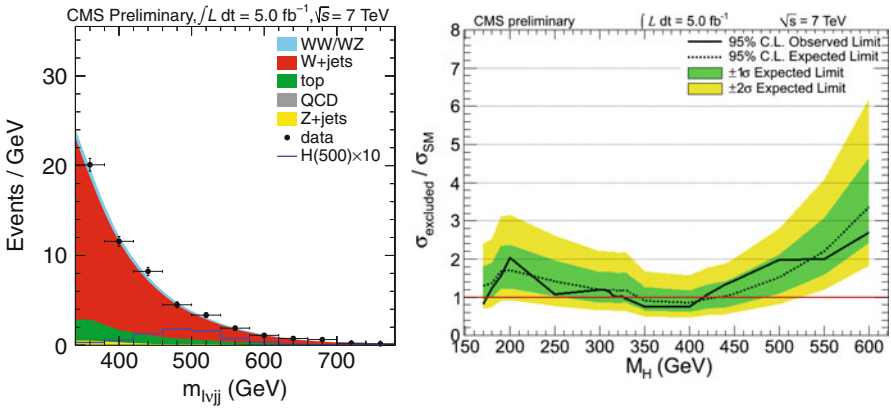

Fig. 18 *Left*: The observed mass spectrum in the CMS $H \rightarrow WW \rightarrow \ell\nu qq$ search [7]. *Right*: Limits on the Higgs boson production rate, in units of the SM Higgs boson rate at 7 TeV [42]

The background has contributions from real di-boson production, rather like the $\ell\nu\ell\nu$ search, but it is swamped by a much larger component from W plus jets, an to a lesser extent by top quarks. Modelling this in detail is not practical, but, as can be seen in Fig. 18, the background is monotonically falling with mass (after a kinematic turn on) and the search is done rather like $H \rightarrow \gamma\gamma$, by looking for a peak on a smooth background.

The CMS results, shown in Fig. 18 (right) show sensitivity to a cross-section which matches that of the SM around 400 GeV but is otherwise weaker. Despite the large signal rate, the very large W plus jets background makes identifying any potential signal difficult.

6.4 Higgs Boson Decay to $b\bar{b}$

The dominant decay mode by rate for Higgs bosons masses below 135 GeV is to a pair of bottom quarks, but searches are complicated by the enormous LHC b-quark production rates in other processes. The situation appears to be most promising in the associated production modes, where the vector boson can also be used to provide a trigger. $WH \rightarrow l\nu bb$, $ZH \rightarrow llbb$ and $ZH \rightarrow \nu\nu bb$ modes all contribute at similar levels.

There are still very large backgrounds from top quarks, VV and especially Vbb production, which are to some extent controlled by analysing the data as a function of the p_T of the Higgs boson candidate. As the signal processes are s-channel, they have harder p_T distributions than the background [61], but the low p_T regions are important to measure the background contributions.

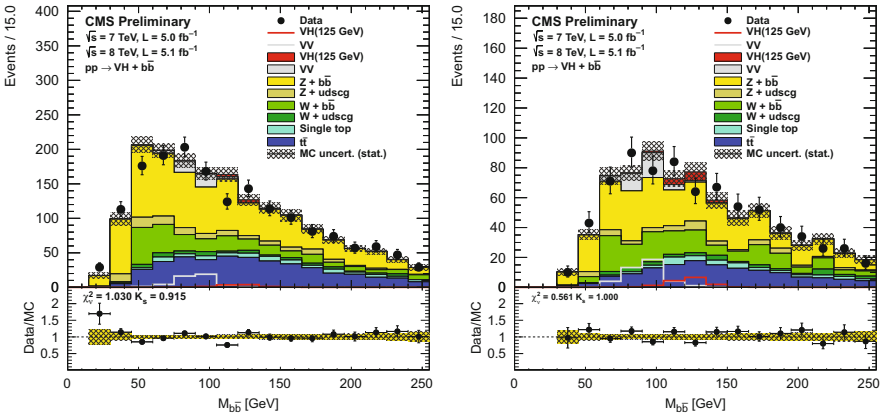

Fig. 19 Mass spectra found by CMS in the VH, $H \rightarrow b\bar{b}$ search [63], separated by candidate boson p_T into low p_T (*left*) and high p_T (*right*) searches

ATLAS has produced results only from the 7 TeV data [62] while CMS analysed the full dataset for the ICHEP meeting [63]. The analysis is done in many divisions of leptons identified and by p_T of the candidate, but in Fig. 19 these are accumulated into two distributions with low and high p_T.

The CMS search has an expected limit around 1.6 times the SM rate at 125 GeV, but the observed limit is somewhat worse, 2.1 times the SM rate, which is approximately the effect expected if there were a signal present. However, the sensitivity is somewhat marginal and it is of great importance to improve these measurements in order to constrain what is expected to be the dominant Higgs boson branching ratio.

6.5 *Higgs Boson Decay to $\tau\,\tau$*

The tau pair searches, similar to those used in the MSSM and mentioned in Sect. 5.1 are also optimised to look for a SM Higgs boson. In this case, partially because the τ can be used as a trigger, the VBF production mode is an important mechanism giving adequate production rate with sufficient rejection of backgrounds, but the boosted and vector boson associated modes also contribute. Unlike the MSSM case, production in association with a b-quark has a very low rate in the SM and is not considered.

ATLAS only analysed the 7 TeV 2011 data [64], while the CMS collaboration used all data up to ICHEP [65]. This search used $e\mu$, $\mu\mu$, $\mu\tau_h$ and $e\tau_h$ tau pair decay modes and analysed the production in terms of VBF, boosted and no-jet signatures.

The mass of an object decaying to τ pairs cannot be completely reconstructed on an event-by-event basis because the τ decays always involve the release of at least

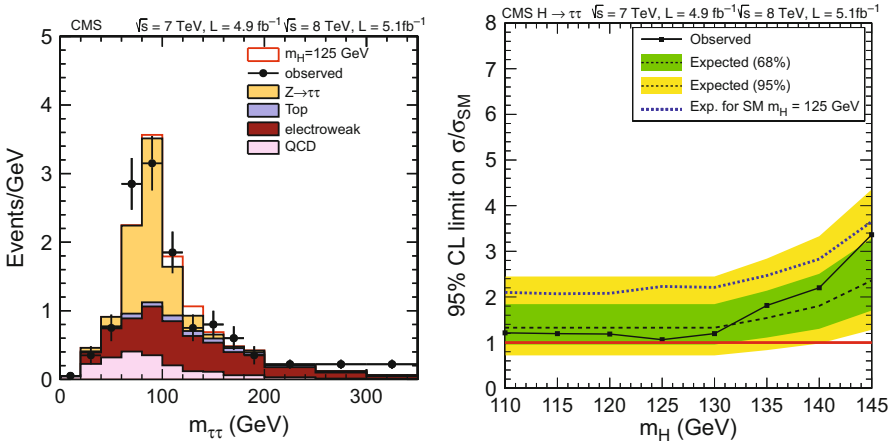

Fig. 20 *Left*: The mass spectrum from CMS VBF $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ search [7]. *Right*: The limits expected and observed in the $H \rightarrow \tau\tau$ search as a function of m_H in units of the Standard Model cross-section

one neutrino each. Various approaches are used to estimate the mass; the simplest is the 'collinear' approximation which is the assumption that the visible decay products are going in the same direction as the original τs. Given that assumption and the missing transverse energy vector (with its uncertainty), the energies of the original τ pairs can also be calculated. Better techniques make use of the measured momentum as well to improve the estimate. The example of the $\mu\tau_h$ channel in the VBF production mode is shown in Fig. 20.

The CMS search was expected to set a limit 1.3 times the SM Higgs boson cross-section, while the observed data actually set a tighter limit, and come close to excluding a SM Higgs boson in this channel. (Note that ATLAS and CMS results released after the School appear more consistent with the expectations including a signal.)

6.6 Combined Search Results

The discovery of the 'new boson' came not from a single channel but from a combined analysis of all the relevant decay modes. But what does relevant mean? The combination relies upon the theoretically expected branching ratios and production cross-sections to build a combined likelihood ratio where every search enters weighted appropriately. In practice all the modes discussed above were combined, but most of the sensitivity came from WW, ZZ and $\gamma\gamma$. In the discovery publications ATLAS only presented these three from the 2012 data, while CMS also included bb and $\tau\tau$.

Table 1 The channels used by each of the experiments, along with the amount of data contributing to them and the mass range for which they produce results

	ATLAS			CMS		
	Luminosity, fb^{-1}		Mass range	Luminosity, fb^{-1}		Mass range
	7 TeV	8 TeV	GeV	7 TeV	8 TeV	GeV
$\gamma\gamma$	4.8	5.9	110–150	5.1	5.3	110–150
$\tau\tau$	4.6	–	–	4.9	5.1	110–145
$\text{b}\bar{\text{b}}$	4.7	–	110–130	5.0	5.1	110–135
$\text{WW} \to \ell\nu\ell\nu$	4.7	5.8	110–240	4.9	5.1	110–600
$\text{WW} \to \ell\nu\text{qq}$	1.04		240–600	–	–	
$\text{ZZ} \to \ell\ell\ell\ell$	4.8	5.9	110–600	5.1	5.3	110–600
$\text{ZZ} \to \ell\ell\nu\nu$	4.7	–	200–600	5.0	5.0	200–600
$\text{ZZ} \to \ell\ell\text{qq}$	4.7	–	200–600	4.6	–	130–600

The decay width of the SM Higgs boson into the different final states depends upon its mass, but it is accurately predicted. It is therefore ideally suited to a combination approach where all the disparate channels are considered together. This is especially true because each channel has its own region of applicability, as shown in Table 1. Most attention has focussed on the low mass region, but at high masses CMS has used the $\text{WW} \to \ell\nu\ell\nu$ decay mode while ATLAS has presented results on $\text{WW} \to \ell\nu\text{qq}$.

In many respects the two experiments are very similar. The sensitivity is largely from the $\gamma\gamma$ search for Higgs boson masses below about 120 GeV, but then the $\text{WW} \to \ell\nu\ell\nu$ search dominates to 200 GeV with $\text{ZZ} \to \ell\ell\ell\ell$ taking a major role beyond. For Higgs boson masses above about 300 GeV the $\text{ZZ} \to \ell\ell\nu\nu$ search has the greatest power. One striking fact is that the $\text{WW} \to \ell\nu\ell\nu$ search by CMS at high mass is important to very high masses and is not even considered by ATLAS.

Both collaborations have produced combinations of their own results, and the p-values, the probability of getting such an excess from background, are shown in Fig. 21.

The evidence for a new particle at 125–126 GeV is striking. Both experiments have local probabilities below 1 in a million. The look elsewhere effect reduces these somewhat. CMS calculates that the significance of such an excess anywhere in the mass range 110–145 GeV is 4.5σ, to be compared with the 5σ local significance. This rather modest reduction is grossly outweighed by the fact that the two experiments see excesses at compatible masses. There is therefore no reasonable doubt that a new particle has been observed.

The $\gamma\gamma$ and $\ell\ell\ell\ell$ channels have a mass resolution of order of 2 GeV: much better than the other modes used. They therefore give independent accurate mass measurements. Figure 22 shows 2D fits to the signal strength and mass in the high resolution channels of ATLAS and CMS. ATLAS also shows the WW decay mode, where the observed data has relatively low sensitivity to the Higgs boson mass, but

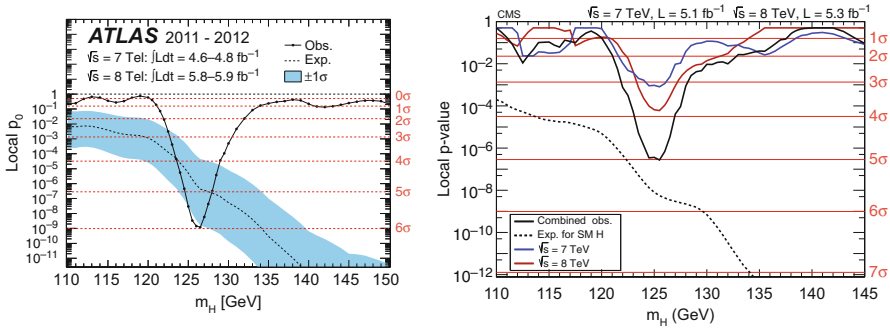

Fig. 21 The p-value of the observation by ATLAS [6] (*left*) and CMS [7] (*right*) in the combination of Standard Model search channels. The ATLAS plot shows the sensitivities if the individual search channels, while the CMS version is divided into the 2 years data. Both by mode and by period the results are distributed roughly in accordance with expectations

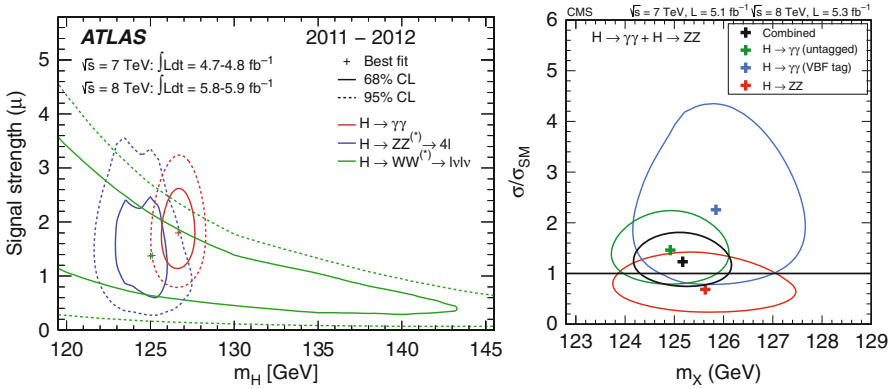

Fig. 22 The 68 % (*solid*) contours for the mass and production rate for the new boson as observed in three modes in ATLAS [6] (*left*) and CMS [7] (*right*). Both experiments show the ZZ and $\gamma\gamma$ modes. CMS has chosen to divide the latter into events which have, or have not, been tagged as VBF candidates. ATLAS in addition shows the WW mode which has comparitively poor mass resolution, but the expected rate is sensitive to the true Higgs boson mass

the expected rate climbs rapidly with increasing m_H. This explains the observed fall in the ratio of observed strength to the SM expectation.

The five high resolution searches all show broadly compatible masses and signal strengths. Thus it seems reasonable to interpret the various excesses as being due to one particle. Given this assumption we can compare its decay rate in various modes to the expectations for the SM Higgs boson, and this is done in Fig. 23.

The observed rates are always within two sigma of the expectations for a Standard Model Higgs boson, while all ten decay modes have some sort of excess compared with the background expectation. The significance of the fermionic modes is very slim; they are perfectly compatible with either hypothesis. It is

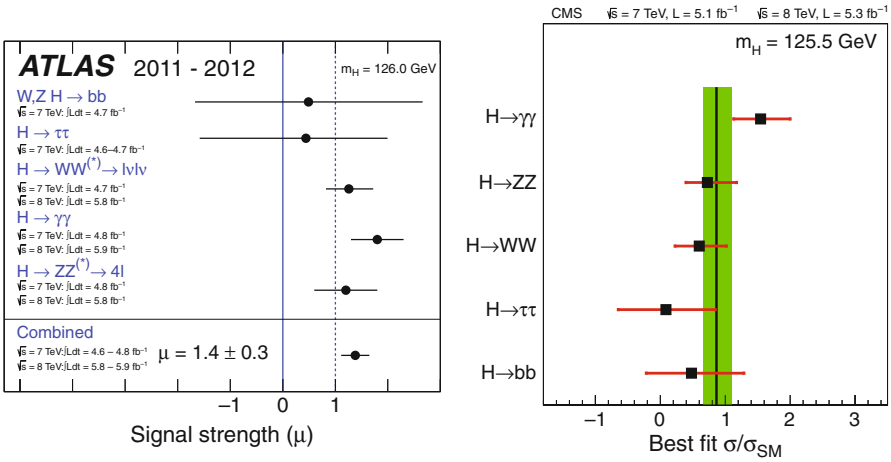

Fig. 23 The measured signal strengths in various Higgs boson decay modes, where by definition a signal strength of 1 denotes agreement with the Standard Model prediction, for ATLAS [6] (*left*) and CMS [7] (*right*)

interesting to note the small excess in the $\gamma\gamma$ mode over the expectation for the SM Higgs which is seen in both experiments. However, at present all the data are completely consistent with reasonable fluctuations around the SM prediction.

7 Summary and Outlook

The first major discovery of the LHC programme is a particle that matched the requirements of the Higgs boson search. At present many of its properties remain obscure, but it is clearly an electrically neutral boson. Furthermore, the decay to $\gamma\gamma$ demonstrates that it cannot have spin 1, unlike the vector bosons already discovered, so it is already clear that this is something radically new. The relative rates of ZZ and $\gamma\gamma$ indicate that it decays ten times as often into weak boson pairs as photons, which is strange given the very large kinematic suppression of the former. This means that it must couple directly to them; i.e. the new particle must carry weak hypercharge. These features make it an excellent candidate for a Higgs boson.

The next steps must be verifying that the observed particle has the correct spin and parity quantum numbers for a Higgs boson, 0^+, and then testing the couplings through measuring the production an decay rates in as many modes as possible. There are many possible implementations of the Higgs mechanism, which usually require more than one Higgs boson. The two Higgs doublet model, and particularly the 'Type-II' variant [19] favoured by supersymmetry, are obvious possible candidates, but nothing should be taken for granted.

The spin and parity can be assessed by studying the angular distributions of the Higgs boson decay products. The three major modes contributing so far – $\gamma\gamma$, WW, and ZZ – all carry significant information, but there are complications. If the new boson is spin two then there is a tensor structure in its production that cannot be predicted without a full model. It is unlikely to be possible to exclude all options with the Run 1 data, although 20 fb^{-1} of data should allow better than 3σ sensitivity to some. For parity analyses a similar sensitivity should hold between 0^+ and 0^-, but the bosonic decays project out the 0^+ component and so are not sensitive to a mixed state. The fermionic modes can address this, but are unlikely to be sensitive without considerably more data. The idea of a mixed CP state is considered in the MSSM with CP violation in the Higgs sector [66].

The major production mode of the Higgs boson, gluon fusion with a top quark loop, would be affected by the presence of other coloured heavy particles. For example a fourth chiral generation of fermions would increase the production rate by a factor 4–9, and is now rather disfavoured. The influence of the MSSM stop squark would be more subtle but would also change the cross-section. The significance of the observation at over five-sigma in only 10 fb^{-1} of LHC data strongly suggests that experimental errors on the total production rate of order 20 % or better should be achievable from Run 1 data, and maybe 10 % by combining experiments. The interpretation will then be limited by the difficulty of these cross-section calculations.

The coupling of the Higgs boson to weak bosons is predicted rather well, and given the electroweak fit precision on the custodial symmetry, is well established. Thus it would be a major surprise if the WW and ZZ rates did not fit expectations. To be more precise, the ratio of these rates, which removes the production cross-section uncertainty, is a good test. The decay to photons, however, is a loop diagram and therefore would be changed by the presence of heavy charged objects, such as the status of supersymmetry. Thus it will be very interesting to measure this rate precisely, probably relative to the weak boson rates.

The fermionic couplings of the Higgs boson are not intrinsic to the breaking of electroweak symmetry, and the masses of the fermions could in principle come from some other mechanism. It is therefore very interesting to test these. At present the constraints are rather weak, but with the full Run 1 data set, and preferably a combination of the two experiments it should be possible to make meaningful statements about bb and $\tau\tau$. The top coupling can also be extracted, but only by making assumptions on the total width. Very interesting, but probably needing hundreds or even thousands of inverse femtobarn, is the di-muon decay. The ratio of the leptonic couplings is extremely well predicted in the SM and will make an important test.

If dark matter exists in the form of WIMPs weighing less than half the mass of the Higgs boson then they may well couple to the Higgs boson and therefore give rise to invisible decays. This has not yet been tested experimentally. It is quite challenging in a hadron collider environment but some sensitivity must be possible in the associated production modes. Thus the Higgs boson may shed some light upon dark matter.

The LHC performance was absolutely excellent and it is clear that the LHC run 1 has not only brought us a Higgs-like boson but also allowed significant studies of it. The approved LHC programme, delivering to 300 fb^{-1} at 13–14 TeV, will deliver about 70 times as many Higgs bosons to the experiments as were used for the discovery, and if they can be triggered and analysed despite the ever-harsher pileup then we will be able to test the model significantly. A factor 10 more data at HL-LHC will open a new set of studies, including perhaps the Higgs self-coupling. The Higgs sector is fast becoming a precision part of the Standard Model.

References

1. G. Aad et al., ATLAS Collaboration, JINST **3**, S08003 (2008)
2. S. Chatrchyan et al., CMS Collaboration, JINST **3**, S08004 (2008)
3. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
4. P.W. Higgs, Phys. Lett. **12**, 132 (1964)
5. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
6. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012)
7. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **716**, 30 (2012)
8. R. Bailey (ed.), *Proceedings of the CAS – CERN Accelerator School: RF for Accelerators*, Ebeltoft, CERN-2011-007 (2011)
9. M. Spira, P.M. Zerwas, Lect. Notes Phys. **512**, 161 (1998)
10. J. Goldstone, Nuovo Cim. **19**, 154 (1961)
11. S.L. Glashow, Nucl. Phys. **22**, 579 (1961)
12. A. Salam, *Elementary Particle Theory* (Almqvist and Wiksell, Stockholm, 1968), p. 367
13. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967)
14. G.'t Hooft, M.J.G. Veltman, Nucl. Phys. B **44**, 189 (1972)
15. J.R. Ellis, M.K. Gaillard, D.V. Nanopoulos, Nucl. Phys. B **106**, 292 (1976)
16. K. Kodama et al., DONUT Collaboration, Phys. Lett. B **504**, 218 (2001)
17. S. Dittmaier et al., LHC Higgs cross section working group collaboration (2011). arXiv:1101.0593 [hep-ph]
18. S. Dittmaier et al., LHC Higgs cross section working group collaboration (2012). arXiv:1201.3084 [hep-ph]
19. V.D. Barger, J.L. Hewett, R.J.N. Phillips, Phys. Rev. D **41**, 3421 (1990)
20. G. Aad et al., ATLAS and CMS Collaborations, ATL-PHYS-PUB-2011-011, CERN-CMS-NOTE-2011-005 (2011)
21. G. Cowan, K. Cranmer, E. Gross, O. Vitells, Eur. Phys. J. C **71**, 1554 (2011)
22. A.L. Read, J. Phys. G **28**, 2693 (2002)
23. E. Gross, O. Vitells, Eur. Phys. J. C **70**, 525 (2010)
24. R. Barate et al., LEP Working Group for Higgs boson searches and ALEPH and DELPHI and L3 and OPAL Collaborations, Phys. Lett. B **565**, 61 (2003)
25. M.M. Kado, C.G. Tully, Ann. Rev. Nucl. Part. Sci. **52**, 65 (2002)
26. G. Abbiendi et al., OPAL Collaboration, Eur. Phys. J. C **27**, 311 (2003)
27. J.R. Ellis, G. Ridolfi, F. Zwirner, Phys. Lett. B **257**, 83 (1991)
28. A. Abulencia et al., CDF Collaboration, Phys. Rev. Lett. **97**, 242003 (2006)
29. T. Aaltonen et al., CDF Collaboration, Phys. Rev. Lett. **108**, 151803 (2012)
30. V.M. Abazov et al., D0 Collaboration, Phys. Rev. Lett. **108**, 151804 (2012)
31. Tevatron Electroweak Working Group and CDF and D0 Collaborations, arXiv:1107.5255 [hep-ex] (2011)

32. Tevatron New Physics Higgs Working Group and CDF and D0 Collaborations, arXiv: 1207.0449 [hep-ex] (2012)
33. T. Aaltonen et al., CDF and D0 Collaborations, Phys. Rev. Lett. **109**, 071804 (2012)
34. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-094 (2012)
35. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **713**, 68 (2012)
36. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-094 (2012)
37. S. Chatrchyan et al., CMS Collaboration, JHEP **1207**, 143 (2012)
38. G. Aad et al., ATLAS Collaboration, JHEP **1206**, 039 (2012)
39. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-098 (2012)
40. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-017 (2012)
41. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **718**, 391 (2012)
42. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-021 (2012)
43. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-092 (2012)
44. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-016 (2012)
45. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **717**, 29 (2012)
46. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-023 (2012)
47. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **717**, 70 (2012)
48. S. Chatrchyan et al., CMS Collaboration, JHEP **1204**, 036 (2012)
49. G. Aad et al., ATLAS Collaboration, Phys. Rev. D **86**, 032003 (2012)
50. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **710**, 26 (2012)
51. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-019
52. ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD and LEP Electroweak Working Group and Tevatron Electroweak Working Group and SLD Electroweak and Heavy Flavour Groups Collaborations, arXiv:1012.2367 [hep-ex] (2010)
53. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **710**, 383 (2012)
54. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. Lett. **108**, 111804 (2012)
55. T. Melia, P. Nason, R. Rontsch, G. Zanderighi, JHEP **1111**, 078 (2011)
56. J.M. Campbell, R.K. Ellis, C. Williams, JHEP **1107**, 018 (2011)
57. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2012-091 (2012)
58. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-015 (2012)
59. M. Dittmar, H.K. Dreiner, Phys. Rev. D **55**, 167 (1997). hep-ph/9608317
60. A.J. Barr, B. Gripaios, C.G. Lester, JHEP **0907**, 072 (2009)
61. J.M. Butterworth, A.R. Davison, M. Rubin, G.P. Salam, Phys. Rev. Lett. **100**, 242001 (2008)
62. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **718**, 369 (2012)
63. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-019 (2012)
64. G. Aad et al., ATLAS Collaboration, JHEP **1209**, 070 (2012)
65. S. Chatrchyan et al., CMS Collaboration, CMS-PAS-HIG-12-018 (2012)
66. M.S. Carena, J.R. Ellis, S. Mrenna, A. Pilaftsis, C.E.M. Wagner, Nucl. Phys. B **659**, 145 (2003)

Flavour Physics in the LHC Era

Tim Gershon

Abstract These lectures give a topical review of heavy flavour physics, in particular CP violation and rare decays, from an experimental point of view. They describe the ongoing motivation to study heavy flavour physics in the LHC era, the current status of the field emphasising key results from previous experiments, some selected topics in which new results are expected in the near future, and a brief look at future projects.

1 Introduction

The concept of "flavour physics" was introduced in the 1970s [1]

> The term flavor was first used in particle physics in the context of the quark model of hadrons. It was coined in 1971 by Murray Gell-Mann and his student at the time, Harald Fritzsch, at a Baskin-Robbins ice-cream store in Pasadena. Just as ice cream has both color and flavor so do quarks.

Leptons also come in different flavours, and flavour physics covers the properties of both sets of fermions. Counting the fundamental parameters of the Standard Model (SM), the 3 lepton masses, 6 quark masses and 4 quark mixing (CKM) matrix [2, 3] parameters are related to flavour physics. In case neutrino masses are introduced, the new parameters (at least 3 more masses and 4 more mixing parameters) are also related to flavour physics. This large number of free parameters is behind several of the mysteries of the SM:

- Why are there so many different fermions?
- What is responsible for their organisation into generations/families?
- Why are there 3 generations/families each of quarks and leptons?
- Why are there flavour symmetries?
- What breaks the flavour symmetries?
- What causes matter – antimatter asymmetry?

T. Gershon (✉)
Department of Physics, University of Warwick, Coventry, UK

European Organization for Nuclear Research (CERN), Geneva, Switzerland
e-mail: T.J.Gershon@warwick.ac.uk

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_6

Unfortunately these mysteries will not be answered in these lectures – they are mentioned here simply because it is important to bear in mind their existence. Instead the focus will be on specific topics in the flavour-changing interactions of the charm and beauty quarks,[1] with occasional digressions on related topics.

While our main interest is in the properties of the charm and beauty quarks, due to the strong interaction, experimental studies must be performed using one or more of the many different charmed or beautiful hadrons. These can decay to an even larger multitude of different final states, making learning the names of all the hadrons a big challenge for flavour physicists. Moreover, hadronic effects can often obscure the underlying dynamics. Nevertheless, it is the hadronisation that results in the very rich phenomenology that will be discussed, so one should bear in mind that [4]

> The strong interaction can be seen either as the "unsung hero" or the "villain" in the story of quark flavour physics.

2 Motivation to Study Heavy Flavour Physics in the LHC Era

There are two main motivations for ongoing experimental investigations into heavy flavour physics: (i) CP violation and its connection to the matter-antimatter asymmetry of the Universe; (ii) discovery potential far beyond the energy frontier via searches for rare or SM forbidden processes. These will be discussed in turn below.

First let us consider one of the mysteries listed above (What breaks the flavour symmetries?) to see how it is connected to these motivations. In the SM, the vacuum expectation value of the Higgs field breaks the electroweak symmetry. Fermion masses arise from the Yukawa couplings of the quarks and charged leptons to the Higgs field, and the CKM matrix arises from the relative misalignment of the Yukawa matrices for the up- and down-type quarks. Consequently, the only flavour-changing interactions are the charged current weak interactions. This means that there are no flavour-changing neutral currents (the GIM mechanism [5]), a feature of the SM which is not generically true in most extended theories. Flavour-changing processes provide sensitive tests of this prediction; as an example, many new physics (NP) models induce contributions to the $\mu \to e\gamma$ transition at levels close to (or even above!) the current experimental limit, recently made more restrictive by the MEG experiment [6], $\mathscr{B}(\mu^+ \to e^+\gamma) < 5.7 \times 10^{-13}$ at 90 % confidence level (CL). Improved experimental reach in this and related charged lepton flavour violation searches therefore provides interesting and unique NP discovery potential (for a review, see, e.g., Ref. [7]).

[1] It is one of the peculiarities of our field that "heavy flavour physics" does not include discussion of the heaviest flavoured particle, the top quark.

2.1 *CP* Violation

As mentioned above, the CKM matrix arises from the relative misalignment of the Yukawa matrices for the up- and down-type quarks:

$$V_{CKM} = U_u U_d^\dagger , \tag{1}$$

where U_u and U_d diagonalise the up- and down-type quark mass matrices respectively. Hence, V_{CKM} is a 3×3 complex unitary matrix. Such a matrix is in general described by 9 (real) parameters, but 5 can be absorbed as unobservable phase differences between the quark fields. This leaves 4 parameters, of which 3 can be expressed as Euler mixing angles, but the fourth makes the CKM matrix complex – and hence the weak interaction couplings differ for quarks and antiquarks, i.e. CP violation arises.

The expression "CP violation" refers to the violation of the symmetry of the combined C and P operators, which replace particle with antiparticle (charge conjugation) and invert all spatial co-ordinates (parity) respectively. Therefore CP violation provides absolute discrimination between particle and antiparticle: one cannot simply swap the definition of which is called "particle" with a simultaneous redefinition of left and right.[2] There is a third discrete symmetry, time reversal (T), and it is important to note that there is a theorem that states that CPT must be conserved in any locally Lorentz invariant quantum field theory [11]. Therefore, under rather reasonable assumptions, an observation of CP violation corresponds to an observation of T violation, and vice versa. Nonetheless, it remains of interest to establish T violation without assumptions regarding other symmetries [12, 13].

The four parameters of the CKM matrix can be expressed in many different ways, but two popular choices are the Chau-Keung (PDG) parametrisation – $(\theta_{12}, \theta_{13}, \theta_{23}, \delta)$ [14] – and the Wolfenstein parametrisation – (λ, A, ρ, η) [15]. In both cases a single parameter (δ or η) is responsible for all CP violation. This encapsulates the predictivity that makes the CKM theory such a remarkable success: it describes a vast range of phenomena at many different energy scales, from nuclear beta transitions to single top quark production, all by only four parameters (plus hadronic effects).

Let us digress a little into history. In 1964, CP violation was discovered in the kaon system [16], but it was not until 1973 that Kobayashi and Maskawa proposed that the effect originated from the existence of three quark families [3]. On a shorter time-scale, in 1967 Sakharov noted that CP violation was one of three conditions necessary for the evolution of a matter-dominated universe, from a symmetric initial state [17]:

[2] The importance of CP violation in this regard was noted by Landau [8] following the observation of parity violation [9, 10].

1. Baryon number violation,
2. C and CP violation,
3. Thermal inequilibrium.

This observation evokes the prescient concluding words of Dirac's 1933 Nobel lecture, discussing his successful prediction of the existence of antimatter, in the form of the positron [18]:

> If we accept the view of complete symmetry between positive and negative electric charge so far as concerns the fundamental laws of Nature, we must regard it rather as an accident that the Earth (and presumably the whole solar system), contains a preponderance of negative electrons and positive protons. It is quite possible that for some of the stars it is the other way about, these stars being built up mainly of positrons and negative protons. In fact, there may be half the stars of each kind. The two kinds of stars would both show exactly the same spectra, and there would be no way of distinguishing them by present astronomical methods.

Dirac was not aware of the existence of CP violation, that breaks the *complete symmetry* of the laws of Nature. Moreover, modern astronomical methods do allow to search for antimatter dominated regions of the Universe, and none have been observed (though searches, for example by the PAMELA and AMS experiments, are ongoing). Therefore, CP violation appears to play a crucial role in the early Universe.

We can illustrate this with a simple exercise. Suppose we start with equal amounts of matter (X) and antimatter ($\bar{X}$). The matter X decays to final state A (with baryon number N_A) with probability p and to final state B (baryon number N_B) with probability $(1-p)$. The antimatter, $\bar{X}$, decays to final state $\bar{A}$ (with baryon number $-N_A$) with probability $\bar{p}$ and final state $\bar{B}$ (baryon number $-N_B$) with probability $(1-\bar{p})$. The resulting baryon asymmetry is

$$\Delta N_{\rm tot} = N_A p + N_B(1-p) - N_A\bar{p} - N_B(1-\bar{p}) = (p-\bar{p})(N_A - N_B)\,.$$

So clearly $\Delta N_{\rm tot} \neq 0$ requires both $p \neq \bar{p}$ and $N_A \neq N_B$, i.e. both CP violation and baryon number violation.

It is natural to next ask whether the magnitude of the baryon asymmetry of the Universe could be caused by the CP violation in the CKM matrix. The baryon asymmetry can be quantified relative to the number of photons in the Universe,

$$\Delta N_B/N_\gamma = (N(\text{baryon}) - N(\text{antibaryon}))/N_\gamma \sim 10^{-10}\,.$$

This can be compared to a dimensionless and parametrisation invariant measure of the amount of CP violation in the SM, $J \times P_u \times P_d / M^{12}$, where

- $J = \cos(\theta_{12})\cos(\theta_{23})\cos^2(\theta_{13})\sin(\theta_{12})\sin(\theta_{23})\sin(\theta_{13})\sin(\delta)$,
- $P_u = (m_t^2 - m_c^2)(m_t^2 - m_u^2)(m_c^2 - m_u^2)$,
- $P_d = (m_b^2 - m_s^2)(m_b^2 - m_d^2)(m_s^2 - m_d^2)$,
- And M is the relevant scale, which can be taken to be the electroweak scale, $\mathcal{O}(100\,\text{GeV})$.

The parameter J is known as the Jarlskog parameter [19], and is expressed above in terms of the Chau-Keung parameters. Putting all the numbers in, we find a value for the asymmetry of $\sim 10^{-17}$, much below the observed 10^{-10}. This is the origin of the widely accepted statement that the SM CP violation is insufficient to explain the observed baryon asymmetry of the Universe. Note that this occurs primarily not because J is small, but rather because the electroweak mass scale is far above the mass of most of the quarks. Therefore, to explain the baryon asymmetry of the Universe, there must be additional sources of CP violation that occur at high energy scales. There is, however, no guarantee that these are connected to the CP violation that we know about. The new sources may show up in the quark sector via discrepancies with CKM predictions (as will be discussed below), but could equally appear in the lepton sector as CP violation in neutrino oscillations. Or, for that matter, new sources could be flavour-conserving and be found in measurements of electric dipole moments, or could be connected to the Higgs sector, or the gauge sector, or to extra dimensions, or to other NP. In any case, precision measurements of flavour observables are generically sensitive to additions to the SM, and hence are well-motivated.

In this context, it is worth noting the enticing possibility of "leptogenesis", where the baryon asymmetry is created via a lepton asymmetry (see, e.g., Ref. [20] for a review). In the case that neutrinos are Majorana particles – i.e. they are their own antiparticles – the right-handed neutrinos may be very massive, which provides an immediate connection with the needed high energy scale. Experimental investigation of this concept requires the determination of the lepton mixing (PMNS) [21, 22] matrix, and proof whether or not neutrinos are Majorana particles. The recent determination of the neutrino mixing angle θ_{13} [23, 24] provides an important step forward; the next challenges are to establish CP violation in neutrino oscillations and to observe (or limit) neutrinoless double beta decay processes.

2.2 Rare Processes

We have already digressed into history, and we should avoid doing so too much, but it is striking how often NP has shown up at the precision frontier before "direct" discoveries at the energy frontier. Examples include: the GIM mechanism being established before the discovery of charm; CP violation being discovered and the CKM theory developed before the discovery of the bottom and top quarks; the observation of weak neutral currents before the discovery of the Z boson. In particular, loop processes are highly sensitive to potential NP contributions, since SM contributions are suppressed or absent.

As a specific example of this we can consider the loop processes involved in oscillations of neutral flavoured mesons. (Rare decay processes will be discussed in more detail below.) There are four such pseudoscalar particles in nature (K^0, D^0, B^0 and B_s^0) which can oscillate into their antiparticles via both short-distance (dispersive) and long-distance (absorptive) processes, as illustrated in Fig. 1. Representing

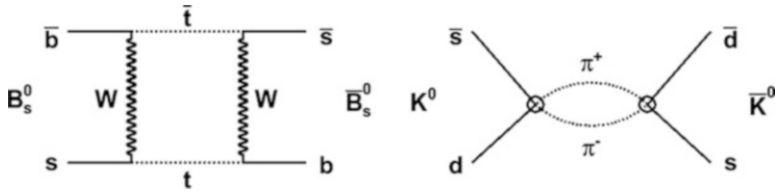

Fig. 1 Illustrative diagrams of (*left*) short-distance (dispersive) processes in B_s^0 mixing; (*right*) long-distance (absorptive) processes in K^0 mixing

such a meson generically by M^0, the evolution of the particle-antiparticle system is given by the time-dependent Schrödinger equation,

$$i\frac{\partial}{\partial t}\begin{pmatrix} M^0 \\ \bar{M}^0 \end{pmatrix} = \left(M - \frac{i}{2}\Gamma\right)\begin{pmatrix} M^0 \\ \bar{M}^0 \end{pmatrix}, \tag{2}$$

where the effective[3] Hamiltonian $H = M - \frac{i}{2}\Gamma$ is written in terms of 2×2 Hermitian matrices M and Γ. Note that the CPT theorem requires that $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$, i.e. that particle and antiparticle have identical masses and lifetimes.

The physical states are eigenstates of the effective Hamiltonian, and are written

$$M_{\rm L,H} = pM^0 \pm q\bar{M}^0, \tag{3}$$

where p and q are complex coefficients that satisfy $|p|^2 + |q|^2 = 1$. Here the subscript labels L and H distinguish the eigenstates by their nature of being lighter or heavier; in some systems the labels S and L are instead used for short-lived and long-lived respectively (the choice depends on the values of the mass and width differences; the labels 1 and 2 are also sometimes used, usually to denote the CP eigenstates). CP is conserved (in mixing) if the physical states correspond to the CP eigenstates, i.e. if $|q/p| = 1$. Solving the Schrödinger equation gives

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}, \tag{4}$$

with eigenvalues given by $\lambda_{\rm L,H} = m_{\rm L,H} - \frac{i}{2}\Gamma_{\rm L,H} = (M_{11} - \frac{i}{2}\Gamma_{11}) \pm (q/p)(M_{12} - \frac{i}{2}\Gamma_{12})$, corresponding to mass and width differences $\Delta m = m_{\rm H} - m_{\rm L}$ and $\Delta\Gamma = \Gamma_{\rm H} - \Gamma_{\rm L}$ given by

$$(\Delta m)^2 - \frac{1}{4}(\Delta\Gamma)^2 = 4(|M_{12}|^2 + \frac{1}{4}|\Gamma_{12}|^2), \tag{5}$$

$$\Delta m\Delta\Gamma = 4\,{\rm Re}(M_{12}\Gamma_{12}^*). \tag{6}$$

[3]The complete Hamiltonian would include all possible final states of decays of M^0 and $\bar{M}^0$.

Table 1 Qualitative expectations and measured values for the neutral meson mixing parameters. Experimental results are taken from Refs. [28–30]. The definition of $a_{\rm sl}$ is given in footnote 6

	Δm ($x = \Delta m/\Gamma$)	$\Delta\Gamma$ ($y = \Delta\Gamma/(2\Gamma)$)	$\lvert q/p\rvert$ ($a_{\rm sl} \approx 1 - \lvert q/p\rvert^2$)
K^0	Large	∼Maximal	Small
	∼500	∼1	$(3.32 \pm 0.06) \times 10^{-3}$
D^0	Small	Small	Small
	$(0.63 \pm 0.19)\,\%$	$(0.75 \pm 0.12)\,\%$	$0.52\,^{+0.19}_{-0.24}$
B^0	Medium	Small	Small
	0.770 ± 0.008	0.008 ± 0.009	-0.0003 ± 0.0021
B_s^0	Large	Medium	Small
	26.49 ± 0.29	0.075 ± 0.010	-0.0109 ± 0.0040

Note that with this notation, which is the same as that of Ref. [25], Δm is positive by definition while $\Delta\Gamma$ can have either sign.[4]

Rather than going into the details of the formalism (which can be found in, e.g., Ref. [27]) let us instead take a simplistic picture.

- The value of Δm depends on the rate of the mixing diagram of Fig. 1(left). This depends on CKM matrix elements, together with various other factors that are either known or (in the case of decay constants and bag parameters) can be calculated using lattice QCD. Moreover for the B mesons, these other factors can be made to cancel in the $\Delta m_d/\Delta m_s$ ratio, such that the measured value of this quantity gives a theoretically clean determination of $|V_{td}/V_{ts}|^2$.
- The value of $\Delta\Gamma$, on the other hand, depends on the widths of decays of the meson and antimeson into common final states (such as CP-eigenstates). Therefore, $\Delta\Gamma$ is large for the K^0 system, where the two pion decay dominates, small for D^0 and B^0 mesons, where the most favoured decays are to flavour-specific or quasi-flavour-specific final states, and intermediate in the B_s^0 system.
- Finally CP violation in mixing tends to zero (i.e. $q/p \approx 1$) if $\arg(\Gamma_{12}/M_{12}) = 0$, $M_{12} \ll \Gamma_{12}$ or $M_{12} \gg \Gamma_{12}$.

This simplistic picture is sufficient to explain qualitatively the experimental values of the mixing parameters given in Table 1. It should be noted that $\Delta\Gamma(B_s^0)$ has become well-measured only very recently (as discussed below), and that the experimental sensitivity for the CP violation parameters in all of the D^0, B^0 and B_s^0 systems is still far from that of the SM prediction, making improved measurements very well motivated.

Thus, neutral meson oscillations are rare processes described by parameters that can be both predicted in the SM and measured experimentally. All measurements

[4]With the definition given, $\Delta\Gamma$ is predicted to be negative for B^0 and B_s^0 mesons in the SM, and hence the sign-flipped definition is often encountered in the literature, e.g. in Ref. [26].

to date are consistent with the SM predictions (though see below). These results can then be used to put limits on non-SM contributions. This can be done within particular models, but the model-independent approach, described in, e.g., Ref. [31] is illustrative. The NP contribution is expressed as a perturbation to the SM Lagrangian,

$$\mathscr{L}_{\rm eff} = \mathscr{L}_{\rm SM} + \Sigma \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathscr{O}_i^d \,({\rm SM\ fields})\,, \tag{7}$$

where the dimension d of higher than 4 has an associated scale Λ and couplings c_i.[5] Given the observables in a given neutral meson system, NP contributions described effectively as four-quark operators ($d = 6$) can be constrained, either by putting bounds on Λ for a fixed value of c_i (typically 1), or by putting bounds on c_i for a fixed value of Λ (typically 1 TeV). In the former case bounds of $\mathscr{O}(100\,{\rm TeV})$ are obtained; in the latter case the bounds can be $\mathscr{O}(10^{-9})$ or below [31], with the strongest (weakest) bounds being in the K^0 (B_s^0) sectors. A similar analysis, but with more up-to-date inputs has been performed in Ref. [32], with results illustrated in Fig. 2. The mixing amplitude, normalised to its SM value, is denoted by Δ, and experimental constraints give $({\rm Re}\Delta, {\rm Im}\Delta)$ consistent with $(1, 0)$ (i.e. with the SM) for both B^0 and B_s^0 systems.

This is a very puzzling situation. Limits on the NP scale give values of at least 100 TeV for generic couplings. But, as discussed elsewhere, we expect NP to appear at the TeV scale to solve the hierarchy problem (and to provide a dark matter candidate, etc.) If NP is indeed at this scale, NP flavour-changing couplings must be small. But why? This is the so-called "new physics flavour problem".

A theoretically attractive solution to this problem, known as minimal flavour violation (MFV) [33], exploits the fact that the SM flavour-changing couplings are also small. Therefore, if there is a perfect alignment of the flavour violation in a NP model with that in the SM, the tension is reduced. The MFV paradigm is highly predictive, stating that there are no new sources of CP violation and also that the correlations between certain observables share their SM pattern (the ratio of branching fractions of $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ being a good example). Therefore, once physics beyond the SM is discovered, it will be an important goal to establish whether or not it is minimally flavour violating. This further underlines that flavour observables carry information about physics at very high scales.

Nonetheless, it must be reiterated that there are several important observables that are not yet well measured, and that could rule out MFV. For example, the bounds on NP scales obtained above (from Ref. [31]) do not include results on CP violation in mixing in the B^0 and B_s^0 sectors. In fact, the D0 collaboration has reported a measurement of an anomalous effect [34] of the inclusive same-sign dimuon asymmetry, which is 3.9σ away from the SM prediction (of very close to

[5] In Eq. (7) it is assumed that the NP modifies the SM operators; more generally extensions to the operator basis are also possible.

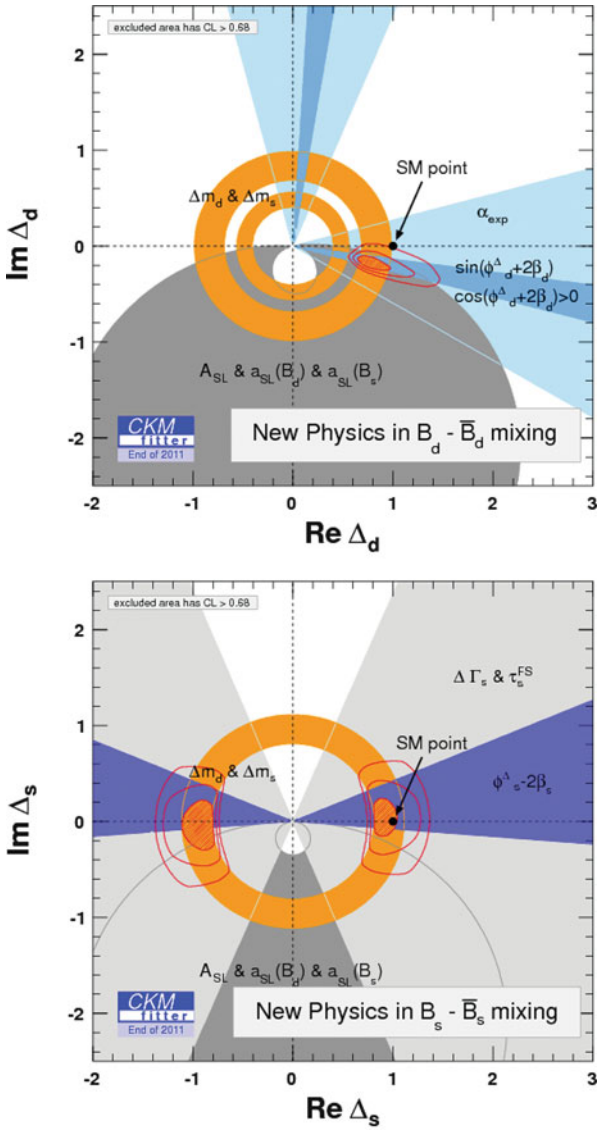

Fig. 2 Constraints on NP contributions in (*top*) B^0 and (*bottom*) B^0_s mixing [32]

zero [35]). This measurement is sensitive to an approximately equal combination of the parameters of CP violation in B^0 and B^0_s mixing, $a^d_{\rm sl}$ and $a^s_{\rm sl}$,[6] however some

[6]The $a_{\rm sl}$ parameters, so named because the asymmetries are measured using semileptonic decays, are related to the p and q parameters by $a_{\rm sl} = (1 - |q/p|^4)/(1 + |q/p|^4)$.

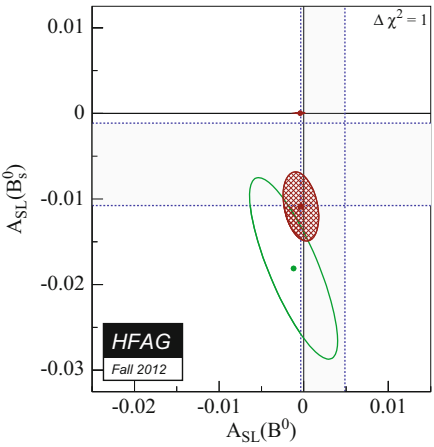

Fig. 3 World average of constraints on the parameters describing CP violation in B^0 and B_s^0 mixing, $a_{\rm sl}^d$ and $a_{\rm sl}^s$. The *green ellipse* comes from the D0 inclusive same-sign dimuon analysis [34]; the *blue shaded bands* give the world average constraints on $a_{\rm sl}^d$ and $a_{\rm sl}^s$ individually; the *red ellipse* is the world average including all constraints [30]

sensitivity to the source of the asymmetry can be obtained by applying additional constraints on the impact parameter to obtain a sample enriched in either oscillated B^0 or B_s^0 candidates. In addition, $a_{\rm sl}^d$ and $a_{\rm sl}^s$ can be measured individually. The latest world average, shown in Fig. 3, gives $a_{\rm sl}^d = -0.0003 \pm 0.0021$, $a_{\rm sl}^s = -0.0109 \pm 0.0040$ [30]. Improved measurements are needed to resolve the situation.

3 Current Experimental Status of Heavy Quark Flavour Physics

3.1 The CKM Matrix and the Unitarity Triangle

Much of the experimental programme in heavy quark flavour physics is devoted to measurements of the parameters of the CKM matrix. As discussed above, the CKM matrix can be written in terms of the Wolfenstein parameters, which exploit the observed hierarchy in the mixing angles:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right), \qquad (8)$$

where the expansion parameter λ is the sine of the Cabibbo angle ($\lambda = \sin\theta_C \approx V_{us}$). It should be noted that although the hierarchy is highly suggestive, there is no underlying reason known for this pattern; moreover, the pattern in the lepton sector is completely different. Note also that at $\mathcal{O}(\lambda^3)$ in the Wolfenstein parametrisation, the complex phase in the CKM matrix enters only in the V_{ub} and V_{td} (top right and bottom left) elements, but this is purely a matter of convention – only relative phases are observable.

The unitarity of the CKM matrix, $V_{\rm CKM}^\dagger V_{\rm CKM} = V_{\rm CKM} V_{\rm CKM}^\dagger = 1$, puts a number of constraints on the magnitudes and relative phases of the elements. Among these relations, one which has been precisely tested is

$$|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1\,, \tag{9}$$

where the measurements of $|V_{ud}|^2$ from, e.g., super-allowed β decays and $|V_{us}|^2$ from leptonic and semileptonic kaon decays are indeed consistent with the prediction to within one part in 10^3 [36].[7]

The unitarity condition also results in six constraints, $\Sigma_i V_{u_i d_j} V^*_{u_i d_k} = \Sigma_i V_{u_j d_i} V^*_{u_k d_i} = 0$ ($u_{i,j,k} \in (u,c,t)$, $d_{i,j,k} \in (d,s,b)$, $j \neq k$), for example

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0\,, \tag{10}$$

which correspond to three complex numbers summing to zero, and hence can be represented as triangles in the complex plane. The triangles have very different shapes, but all of them have the same area, which is given by half of the Jarlskog parameter [19]. The specific traingle relation given in Eq. (10) is particularly useful to visualise the constraints from various different measurements, as shown in the iconic images from the CKMfitter [37] and UTfit [38] collaborations, reproduced in Fig. 4. Conventionally, this "Unitary Triangle" (UT) is rescaled by $V_{cd} V_{cb}^*$ so that by definition the position of the apex is

$$\bar\rho + i\,\bar\eta \equiv -\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}\,, \tag{11}$$

where $(\overline{\rho}, \overline{\eta})$ [39] are related to the Wolfenstein parameters by

$$\rho + i\,\eta \;=\; \frac{\sqrt{1 - A^2\lambda^4}(\bar\rho + i\,\bar\eta)}{\sqrt{1-\lambda^2}\,[1 - A^2\lambda^4(\bar\rho + i\,\bar\eta)]}\,. \tag{12}$$

Two popular naming conventions for the UT angles exist in the literature:

$$\alpha \equiv \phi_2 = \arg\left[-\frac{V_{td} V_{tb}^*}{V_{ud} V_{ub}^*}\right],\quad \beta \equiv \phi_1 = \arg\left[-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*}\right],\quad \gamma \equiv \phi_3 = \arg\left[-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}\right]. \tag{13}$$

[7] The contribution from $|V_{ub}|^2$ is at the level of 10^{-5} and therefore negligible for this test at current precision.

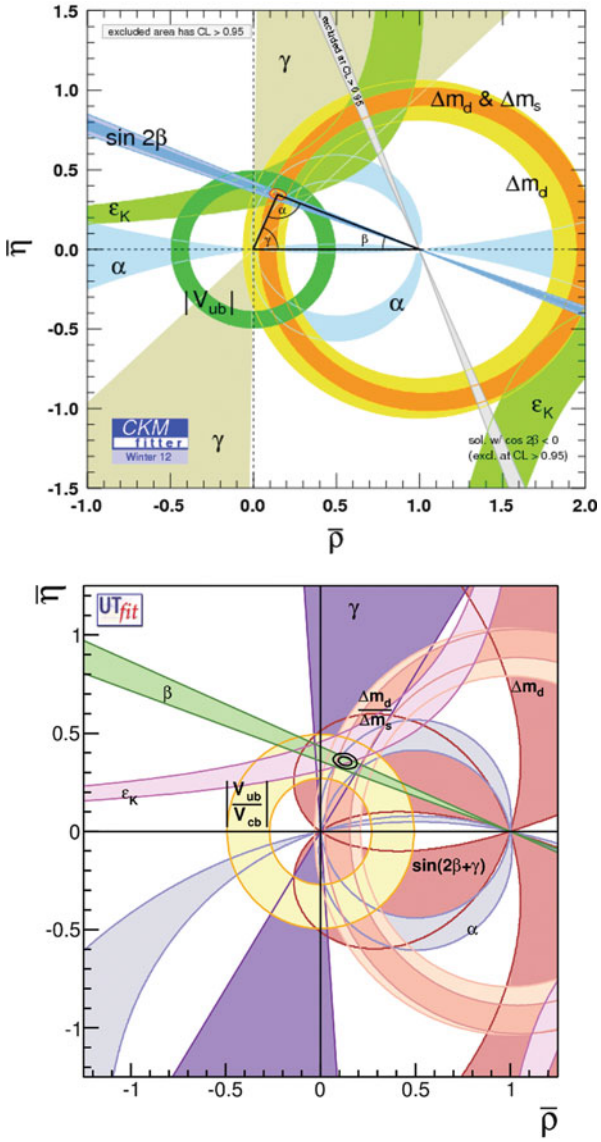

Fig. 4 Constraints on the Unitarity Triangle as compiled by (*top*) CKMfitter [37], (*bottom*) UTfit [38]

The (α, β, γ) set is used in these lectures. The lengths of the sides R_u and R_t of the UT are given by

$$R_u = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right| = \sqrt{\overline{\rho}^2 + \overline{\eta}^2}\,, \qquad R_t = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right| = \sqrt{(1-\overline{\rho})^2 + \overline{\eta}^2}\,. \tag{14}$$

A major achievement of the past decade or so has been to significantly improve the precision of the parameters of the UT. In particular, the primary purpose of the so-called "B factory" experiments, BaBar and Belle, was the determination of $\sin 2\beta$ using $B^0 \to J/\psi\, K_S^0$ (and related modes). This was carried out using completely new experimental techniques to probe CP violation in a very different way to previous experiments in the kaon system. In particular, if we denote the amplitude for a B^0 meson to decay to a particular final state f as A_f, and that for the charge conjugate process as $\bar{A}_{\bar{f}}$, then using the parameters p and q from Eq. (3), we define the parameter $\lambda_f = \frac{q}{p}\frac{\bar{A}_{\bar{f}}}{A_f}$ and the following categories of CP violation in hadronic systems[8]:

1. CP violation in mixing ($|q/p| \neq 1$),
2. CP violation in decay ($\left|\bar{A}_{\bar{f}}/A_f\right| \neq 1$),
3. CP violation in interference between mixing and decay ($\mathrm{Im}\left(\lambda_f\right) \neq 0$).

Additionally, in the literature the concepts of *indirect* and *direct* CP violation are often encountered: the former is where the effect is consistent with originating from a single phase in the mixing amplitude, while the latter cannot be accounted for in such a way. Following this categorisation, CP violation in decay (the only category available to baryons or charged mesons) is direct, while CP violation in mixing and interference can be indirect so long as only one measurement is considered – but if two such measurements give different values, this also establishes direct CP violation.

3.2 *Determination of* $\sin(2\beta)$

The determination of $\sin(2\beta)$ from $B^0 \to J/\psi\, K_S^0$ [40, 41], exploits the fact that some measurements of CP violation in interference between mixing and decay can be cleanly interpreted theoretically, since hadronic factors do not contribute. The full derivation of the decay-time-dependent decay rate of B^0 mesons to a CP eigenstate f is a worthwhile exercise for the reader, and can be found in, e.g., Refs. [42, 43]. The result, for mesons that are known to be either $\overline{B}^0$ or B^0 at time $t = 0$, is

[8] Considering the possibility that CP violation may be observed in the lepton sector as differences of oscillation parameters between neutrinos and antineutrinos (in appearance experiments), it is worth noting that this would be another different category.

$$\Gamma(\overline{B}^0_{\rm phys} \to f(t)) = \frac{e^{-t/\tau(B^0)}}{2\tau(B^0)}\left[1 + S_f \sin(\Delta m t) - C_f \cos(\Delta m t)\right],$$
$$\Gamma(B^0_{\rm phys} \to f(t)) = \frac{e^{-t/\tau(B^0)}}{2\tau(B^0)}\left[1 - S_f \sin(\Delta m t) + C_f \cos(\Delta m t)\right], \quad (15)$$

where

$$S_f = \frac{2\,{\rm Im}(\lambda_f)}{1+|\lambda_f|^2}\,, \text{ and } C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}\,. \quad (16)$$

In these expressions $\Delta\Gamma$ has been assumed to be negligible, as appropriate for the B^0 system. Assuming further $|q/p| = 1$, then for decays dominated by a single amplitude, $C_f = 0$ and $S_f = \sin(\arg(\lambda_f))$, and so for $B^0 \to J/\psi\, K^0_{\rm S}$, $S = \sin(2\beta)$, to a very good approximation.

The experimental challenge for the measurement of $\sin(2\beta)$ then lies in the ability to measure the coefficient of the sinusoidal oscillation of the decay-time-asymmetry. Until recently, the most copious sources of cleanly reconstructed B mesons came from accelerators colliding electrons with positrons at the $\Upsilon(4S)$ resonance (a $b\bar{b}$ bound state just above the threshold for decay into pairs of B mesons). For symmetric colliders, the B mesons are produced at rest, and therefore lifetime measurements are not possible. A boost is necessary, which can be advantageously achieved by making the e^+e^- collisions asymmetric.[9] One strong feature of this approach is that the quantum correlations of the B mesons produced in $\Upsilon(4S)$ decay are retained, so that the decay of one into a final state that tags its flavour (B^0 or $\overline{B}^0$) can be used to set the clock to $t = 0$ and specify the flavour of the other at that time.

The concept of the asymmetric B factory was such a good one that two were built: PEP-II at SLAC, colliding 9.0 GeV e^- with 3.1 GeV e^+, and KEKB at KEK (8.0 GeV e^- on 3.5 GeV e^+). These have achieved world record luminosities, with peak instantaneous luminosities above $10^{34}\,{\rm cm}^{-2}\,{\rm s}^{-1}$, and a combined data sample of over $1\,{\rm ab}^{-1}$, corresponding to over 10^9 $B\bar{B}$ pairs. The detectors (BaBar [44,45] and Belle [46] respectively) share many common features, such as silicon vertex detectors, gas based drift chambers, electromagnetic calorimeters based on Tl-doped CsI crystals, and 1.5 T solenoidal magnetic fields. The main difference is in the technology used to separate kaons from pions: a system based on the detection of internally reflected Cherenkov light for BaBar, and a combination of aerogel Cherenkov counters and a time-of-flight system for Belle.

Through the measurement of $\sin 2\beta$, BaBar [47] and Belle [48] were able to make the first observations of CP violation outside the kaon sector, thus validating the Kobayashi-Maskawa mechanism. The latest (and, excluding upgrades, most

[9] Boosted b hadrons can also be obtained in hadron colliders, as will be discussed below.

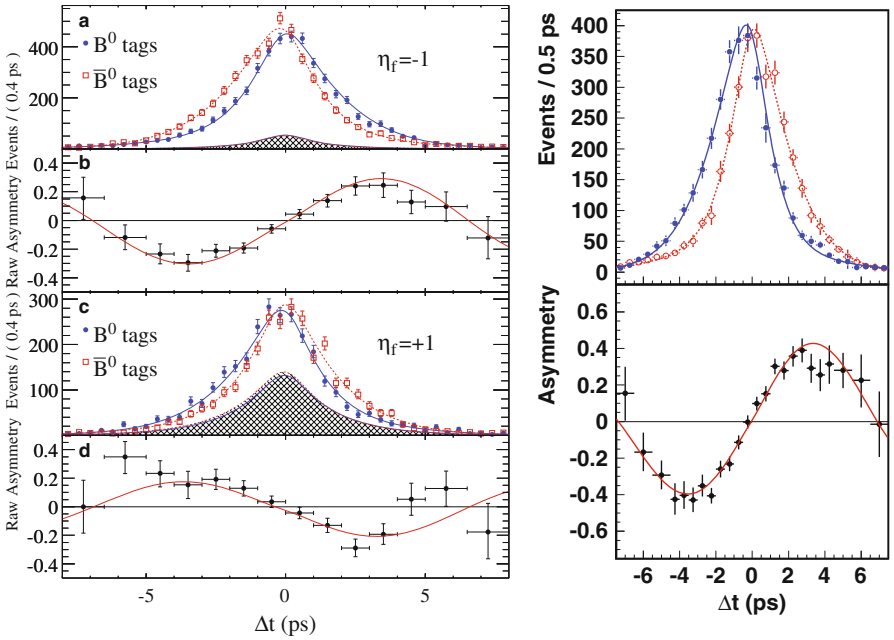

Fig. 5 Results from (*left*) BaBar [49] and (*right*) Belle [50] on the determination of $\sin 2\beta$

likely final) results from BaBar [49] and Belle [50] shown in Fig. 5 give a clear visual confirmation of the large CP violation effect. The world average value, using determinations based on $b \to c\bar{c}s$ transitions, is [30]

$$\sin 2\beta = 0.682 \pm 0.019 \text{ which gives } \beta = (21.5\,^{+0.8}_{-0.7})^\circ . \tag{17}$$

3.3 Determination of α

Additional measurements are needed to over-constrain the UT and thereby test the Standard Model. The angle α can, in principle, be determined in a similar way as β, but using a decay mediated by the $b \to u\bar{u}d$ tree-diagram which carries the relative weak phase γ (since $\pi - (\beta + \gamma) = \alpha$ by definition). However, in any such decay a contribution from the $b \to d$ loop ("penguin") amplitude, which carries a different weak phase, is also possible. This complicates the interpretation of the observables, since $S \neq \sin(2\alpha)$; on the other hand direct CP violation becomes observable, if the relative strong phase is non-zero. Constraints on α can still be obtained using a channel in which the penguin contribution either can be shown to be small, or can be

corrected for using an isospin analysis [51]. The world average, $\alpha = \left(89.0\,^{+4.4}_{-4.2}\right)^\circ$, is dominated by constraints from the $B^0 \to \rho^+\rho^-$ decay [52,53], which is consistent with having negligible penguin contribution.

3.4 The Sides of the Unitarity Triangle

The lengths of the sides of the UT have also been constrained by various observables. The value of R_t depends on $|V_{td}|$, and can be determined from $b \to d$ transitions such as the rate of B^0 oscillations, i.e. Δm_d, or the branching fraction $B \to \rho\gamma$. In both cases, theoretical uncertainties are reduced if the measurement is performed relative to that for the corresponding $b \to s$ transition. The most precise constraint to date comes from the ratio of Δm_d [30,54,55] and Δm_s [30,56,57] and gives $\left|\frac{V_{td}}{V_{ts}}\right| = 0.211 \pm 0.001 \pm 0.005$, where the first uncertainty is experimental and the second theoretical (originating from lattice QCD calculations).

The value of R_u depends on $|V_{ub}|$ and can be determined from $b \to u$ tree-level transitions. Semileptonic decays allow relatively clean theoretical interpretation,[10] but still require QCD calculations to go from the parton level transition to the observed (semi-hadronic) final state (for a recent review, see Ref. [58]). Two approaches have been pursued: exclusive decays, such as $B^0 \to \pi^- e^+ \nu$, and inclusive decays, $B \to X_u e^+ \nu$. The theory of inclusive decays is based on the operator product expansion (discussed in Sect. 5.7) and would be extremely clean, were it not for the fact that experimentally cuts are needed to remove the more prevalent $b \to c$ transition. Exclusive decays tend to have less background from $b \to c$ processes. The differential branching fractions can be translated in constraints on $|V_{ub}|$ using knowledge of form-factors at the kinematic limit obtained from lattice QCD calculations, together with phenomenological models that extrapolate over the whole phase space. The most precise results use $B \to \pi\ell^+\nu$ decays ($\ell = e, \mu$) [59–61], and give an "exclusive" determination of $|V_{ub}|$ that is, however, in tension with the "inclusive" value [30]:

$$|V_{ub}|_{\text{excl.}} = [3.23 \times (1.00 \pm 0.05 \pm 0.08)] \times 10^{-3}\,,$$
$$|V_{ub}|_{\text{incl.}} = [4.42 \times (1.000 \pm 0.045 \pm 0.034)] \times 10^{-3}\,.$$

where the first uncertainties are experimental and the second theoretical. Since the origin of the discrepancy, which is also seen in determinations of $|V_{cb}|$ from $b \to c\ell\nu$ transitions, is not understood, the uncertainty is scaled to give

$$|V_{ub}|_{\text{avg.}} = [3.95 \times (1.000 \pm 0.096 \pm 0.099)] \times 10^{-3}\,.$$

[10] Fully leptonic decays are even cleaner theoretically, but are experimentally scarce. Such modes will be discussed below.

The results on β, α, R_t and R_u are the most constraining inputs to the CKM fits shown in Fig. 4 [37,38]. While the results are all consistent with the Standard Model prediction of a single source of CP violation, there are some tensions that deserve further investigation. Moreover, there are still certain important observables where large NP contributions are possible, as will be discussed in more detail below.

4 Flavour Physics at Hadron Colliders

Results from the B factory experiments provided enormous progress in the understanding of heavy flavour physics (only a very brief selection has been discussed above). Nonetheless, many results remain statistically limited, and the B_s^0 sector is relatively unexplored. To progress further, it is necessary to have a copious source of production of all flavours of b hadron. As shown in Table 2, high energy hadron colliders satisfy these criteria, but present significant experimental challenges: to be able to identify the decays of interest from the high multiplicity environment, and to reject the even more copious rate of minimum bias events.[11]

The LHCb detector [63], shown in Fig. 6, has been designed to meet these challenges. It is in essence a forward spectrometer (covering the acceptance region that optimises its flavour physics capability), with a dipole magnet, a precision silicon vertex detector and strong particle identification capability. Tracks can be identified as different hadron species using information from ring-imaging Cherenkov detectors, while calorimeters and muon detectors enable charged leptons

Table 2 Summary of some relevant properties for b physics in different experimental environments (Adapted from Ref. [62])

	$e^+e^- \to \Upsilon(4S) \to B\bar{B}$ PEP-II, KEKB	$p\bar{p} \to b\bar{b}X$ $(\sqrt{s} = 2\,\text{TeV})$ Tevatron	$pp \to b\bar{b}X$ $(\sqrt{s} = 14\,\text{TeV})$ LHC
Production cross-section	1 nb	$\sim$100 μb	$\sim$500 μb
Typical $b\bar{b}$ rate	10 Hz	$\sim$100 kHz	$\sim$500 kHz
Pile-up	0	1.7	0.5–20
b hadron mixture	B^+B^- (50 %), $B^0\overline{B}^0$ (50 %)	B^+ (40 %), B^0 (40 %), B_s^0 (10 %),	
		Λ_b^0 (10 %), others (<1 %)	
b hadron boost	Small ($\beta\gamma \sim 0.5$)	Large ($\beta\gamma \sim 100$)	
Underlying event	$B\bar{B}$ pair alone	Many additional particles	
Production vertex	Not reconstructed	Reconstructed from many tracks	
B^0–$\overline{B}^0$ pair production	Coherent (from $\Upsilon(4S)$ decay)	Incoherent	
Flavour tagging power	$\epsilon D^2 \sim 30\,\%$	$\epsilon D^2 \sim 5\,\%$	

[11] Experiments at e^+e^- machines also have to reject effectively backgrounds from QED processes, but this can be done at trigger level with simple requirements.

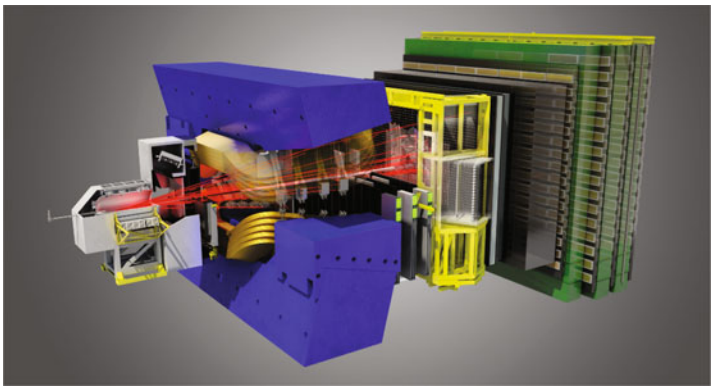

Fig. 6 The LHCb detector [63]

to be distinguished and also provide trigger signals. The trigger system [64] uses these hardware level signals to reduce the rate from the maximum LHC bunch-crossing rate of 40 MHz to the 1 MHz rate at which the detector can be read out. A software trigger then searches for inclusive signatures of b-hadron decays such as high-$p_{\rm T}$ signals and displaced vertices, and also performs reconstruction of several exclusive b and c decay channels, in order to further reduce the rate to a level that can be written to offline data storage (3 kHz in 2011, 5 kHz in 2012).

During the LHC run, the detector operated with data taking efficiency above 90 %, with instantaneous luminosity around $3\,(4) \times 10^{32}\,{\rm cm}^{-2}\,{\rm s}^{-1}$ recording data samples of $1\,(2)\,{\rm fb}^{-1}$ at $\sqrt{s} = 7\,(8)\,{\rm TeV}$ in 2011 (2012).[12] The luminosity is less than that delivered to ATLAS and CMS, since the experimental design requires low pile-up, i.e. a low number of pp collisions per bunch-crossing. However, this allows the luminosity to be "levelled" and remain at a constant value throughout the LHC fill, providing very stable data taking-conditions.[13] In addition, the polarity of the dipole magnet is reversed regularly, which enables cancellation of detector asymmetries in various measurements.

In addition to LHCb, it must be noted that the "general purpose detectors" ATLAS and CMS at the LHC, and CDF and D0 at the Tevatron, have capability to study flavour physics. For most of these experiments, their programme is, however, restricted to decay modes triggered by high $p_{\rm T}$ muons, but CDF benefited from a two-track trigger [65] that enabled a broader range of measurements to be performed.

[12] Note that these values already exceed the LHCb design luminosity of $2 \times 10^{32}\,{\rm cm}^{-2}\,{\rm s}^{-1}$.

[13] Similar stability was achieved at e^+e^- colliders by a completely different method referred to as trickle (or continuous) injection.

4.1 Heavy Flavour Production and Spectroscopy

The capabilities of the different experiments can be demonstrated from the measurements of production cross-sections that have been performed by each. Most have studied J/ψ production (e.g. Refs. [66–71]) as well as b hadron production using decay modes involving muons or J/ψ mesons [72–77]. However, only CDF and LHCb have been able to study prompt charm production [78, 79].[14] The cross-sections measured confirm the theoretical predictions, and enable the values for integrated luminosity to be translated into more easily comprehensible terms. For example, with $1\,\text{fb}^{-1}$ recorded at $\sqrt{s} = 7\,\text{TeV}$, and the measured $b\bar{b}$ production cross-section [77, 80], it is easily shown that over 10^{11} $b\bar{b}$ quark pairs have been produced in the LHCb acceptance. This compares to the combined BaBar and Belle data sample of $\sim 10^9$ $B\bar{B}$ meson pairs. Consequently, for any channel where the efficiency, including effects from reconstruction, trigger and offline selection, is not too small, LHCb has the world's largest data sample. This further emphasises the need for an excellent trigger in order to perform flavour physics at hadron colliders.

Production measurements such as those mentioned above test QCD models, and are important (and highly-cited) results. However, since they are not within the remit of flavour-changing interactions of the charm and beauty quarks, they will not be discussed further here. Nonetheless, a brief digression into studies of another aspect of QCD, that of spectroscopy, will be worthwhile. This covers the study of properties of hadronic states such as lifetimes, masses, decay channels and quantum numbers, and also the discoveries of new states. Indeed, some of the most highly-cited papers from recent flavour physics experiments relate to such topics, including the discovery of the $X(3872)$ particle by Belle [81] and of the D_{sJ} states by BaBar [82] and CLEO [83]. The first new particles discovered at the LHC, prior to the Higgs boson, were hadronic states [84–86]. More recently, significant progress has been made in understanding the nature of the $X(3872)$ [87]. New results are eagerly anticipated in several related areas, for example to clarify the situation regarding the existence of charged charmonium-like states, claimed by Belle [88–90] but not confirmed by BaBar [91, 92], which would be smoking gun signatures for an exotic hadronic nature.[15] Recent claims of charged bottomonium-like states by Belle [95, 96] seem to strengthen the case that such exotics can exist in nature, but one should note that history teaches us that not all claimed states turn out to be real [97].

The topic of spectroscopy also provides a useful illustration of the importance of triggering for flavour physics experiments at hadron colliders. In 2008, the BaBar experiment discovered the η_b meson using the process $e^+e^- \to \Upsilon(3S) \to \eta_b\gamma$, where only the photon is reconstructed and the signal is inferred from a peak in

[14] Measurements of charm production and other processes by ALICE are not included in this discussion. Although ALICE can study production at low luminosity, it cannot perform competitive studies of flavour changing processes.

[15] New claims of charged charmonium-like states have recently been made [93, 94].

the photon energy spectrum [98]. The η_b meson is the pseudoscalar $b\bar{b}$ ground state. It is the lightest bottomonium state, so why did it take more than 30 years after the discovery of the $\Upsilon(1S)$ meson [99] (the lightest vector $b\bar{b}$ state) to see it in experiments? In particular, since hadron collisions produce particles with all possible quantum numbers, why was it not observed at, e.g., the Tevatron? The answer lies in the fact that the vector state decays to dimuons, which have a distinctive trigger signature. The dominant decay channels of the η_b are expected to be multibody hadronic final states, which make its observation in a hadronic environment extremely challenging.

5 Key Observables in the LHC Era

5.1 Direct *CP* Violation

As mentioned above, a condition for direct CP violation is $\left|\bar{A}_{\bar{f}}/A_f\right| \neq 1$. In order for this to be realised we need the amplitude to be composed of at least two parts with different weak and strong phases. This is often realised by tree (T) and penguin (P) amplitudes (example diagrams are shown in Fig. 7), so that

$$A_f = |T|\, e^{i(\delta_T - \phi_T)} + |P|\, e^{i(\delta_P - \phi_P)} \text{ and } \bar{A}_{\bar{f}} = |T|\, e^{i(\delta_T + \phi_T)} + |P|\, e^{i(\delta_P + \phi_P)}\,, \quad (18)$$

where the strong (weak) phases $\delta_{T,P}$ ($\phi_{T,P}$) keep the same (change) sign under the CP transformation by definition. The CP asymmetry is defined from the rate difference between the particle involving the quark (D or $\bar{B}$) and that containing the antiquark ($\bar{D}$ or B). Using the definition for B decays, we trivially find

$$A_{CP} = \frac{\left|\bar{A}_{\bar{f}}\right|^2 - \left|A_f\right|^2}{\left|\bar{A}_{\bar{f}}\right|^2 + \left|A_f\right|^2} = \frac{2\,|T|\,|P|\sin(\delta_T - \delta_P)\sin(\phi_T - \phi_P)}{|T|^2 + |P|^2 + 2\,|T|\,|P|\cos(\delta_T - \delta_P)\cos(\phi_T - \phi_P)}\,. \quad (19)$$

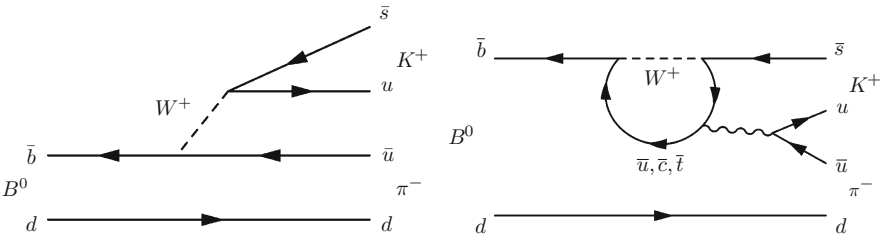

Fig. 7 SM (*left*) tree and (*right*) penguin diagrams for the decays $B^0 \to K^+\pi^-$

Therefore, for large direct CP violation effects to occur, we need $|P/T|$, $\sin(\delta_T - \delta_P)$ and $\sin(\phi_T - \phi_P)$ to all be $\mathcal{O}(1)$.

Charmless B decays, i.e. decays of B mesons to final states that do not contain charm quarks, provide good possibilities for the observation of direct CP violation, since many decays have both tree and penguin contributions with similar levels of CKM suppression. These are of interest to search for NP, since the penguin loop diagrams are sensitive to potential contributions from new particles. An excellent example is $B^0 \to K^+\pi^-$, which provided the first observation of direct CP violation outside the kaon sector, and has a world average value of $A_{CP}(B^0 \to K^+\pi^-) = -0.086 \pm 0.007$ [30, 100–103]. Curiously, the CP violation effect observed in $B^+ \to K^+\pi^0$ decays is rather different: $A_{CP}(B^+ \to K^+\pi^0) = 0.040 \pm 0.021$ [30, 100, 101], although naïvely changing the spectator quark in Fig. 7 suggests that similar values should be expected. This is referred to as the "$K\pi$ puzzle", and could in principle be a hint for NP, though the more mundane explanation of larger than expected QCD corrections cannot be ruled out at present. Several methods are available to test the QCD explanations, which motivate improved measurements of other $K\pi$ modes (in particular, of $A_{CP}(B^0 \to K^0_S\pi^0)$), of similar decay modes with three-body final states ($K\rho, K^*\pi$), and of charmless two-body B^0_s decays. On this last topic, following pioneering work by CDF [103, 104], LHCb has recently reported both the first decay time-dependent analysis of $B^0_s \to K^+K^-$ [105] and the first observation of CP violation in $B^0_s \to K^-\pi^+$ decays [106], which demonstrate good prospects for progress in the coming years.

With regard to three-body decays, it is worth noting that despite hundreds of measurements by the B factories, the significance of the world average in any other charmless B^+ or B^0 decay mode does not exceed 5σ, though channels such as $B^+ \to \eta K^+$ and $B^+ \to \rho^0 K^+$ approach this level. However, very recently, LHCb has demonstrated that large CP violation effects occur in specific regions of the phase space of three-body charmless decays such as $B^+ \to K^+\pi^+\pi^-$ [107–109]. Further study is necessary to quantify the effect and identify its origin.

5.2 The UT Angle γ from $B \to DK$ Decays

The angle γ of the CKM Unitarity Triangle is unique in that it is the only CP-violating parameter that can be measured using only tree-level decays. This makes it a benchmark Standard Model reference point. Improving the precision with which γ is known is one of the primary goals of contemporary flavour physics, and this will only become more important after NP is discovered, since it will be essential to disentangle SM and NP contributions to CP-violating observables.

The phase γ can be determined exploiting the fact that in decays of the type $B \to DK$, the $b \to c\bar{u}s$ and $b \to u\bar{c}s$ amplitudes can interfere if the neutral charmed meson is reconstructed in a final state that is accessible to both D^0 and $\overline{D}^0$ decays. There are many possible such final states, with various

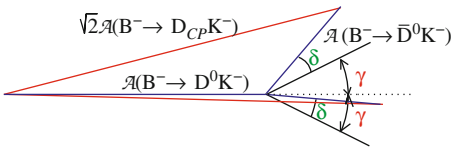

Fig. 8 Illustration of the concept behind the determination of γ using $B^{\pm} \to DK^{\pm}$ decays. For B^- decays the amplitudes add with relative phase $\delta - \gamma$, while for B^+ the relative phase is $\delta + \gamma$. Here the simplest case with D decays to CP eigenstates (such as K^+K^-) is shown, but the method can be extended to any final state accessible to both D^0 and $\overline{D}^0$ decays

experimental advantages and disadvantages. These include CP eigenstates, doubly- or singly-Cabibbo-suppressed decays and multibody final states. Moreover, decays of different b hadrons can all be used to provide constraints on γ. Two particularly interesting approaches are to study decay time-dependent asymmetries of $B_s^0 \to D_s^{\mp} K^{\pm}$ decays [110] and to study the Dalitz plot (i.e. phase-space) dependent asymmetries in $B^0 \to DK^+\pi^-$ decays [111, 112]. First results from LHCb show promising potential for these decays [113, 114]. All such measurements will help to improve the overall precision in a combined fit.

The basic concept behind the method is illustrated in Fig. 8 for $B^- \to D_{CP} K^-$ decays. It must be emphasised that due to the absence of loop contributions to the decay it is extremely clean theoretically [115]. This, and the abundance of different final states accessible, means that all parameters can be determined from data. The relevant parameters are the weak phase γ, an associated strong (CP conserving) phase difference between the $b \to c\bar{u}s$ and $b \to u\bar{c}s$ decay amplitudes, labelled δ_B, and the ratio of their magnitudes, r_B. The small value of $r_B(B^- \to DK^-) \sim 10\,\%$ means that large event samples are necessary to obtain good constraints on γ, and only recently has the first 5σ observation of CP violation in $B \to DK$ decays been achieved [116]. Larger values of r_B are expected in $B^0 \to DK^{*0}$ and $B_s^0 \to D_s^{\mp} K^{\pm}$ decays, but until now the samples available in these channels have not been sufficient to give meaningful constraints on γ. The available measurements use $B^{(*)-} \to D^{(*)} K^{(*)-}$ decays, with the latest combinations from each experiment giving (BaBar) $\gamma = (69^{+17}_{-16})^\circ$ [117], (Belle) $\gamma = (68^{+15}_{-14})^\circ$ [118] and (LHCb) $\gamma = (71^{+15}_{-16})^\circ$ [119]. Significant progress in this area is anticipated from LHCb in the coming years.[16]

5.3 Mixing and CP Violation in the B_s^0 System

A complete analysis of the time-dependent decay rates of neutral B mesons must also take into account the lifetime difference between the eigenstates of the effective Hamiltonian, denoted by $\Delta\Gamma$. This is particularly important in the B_s^0 system, since

[16] Updates using more data have already started to appear from LHCb [120, 121].

the value of $\Delta\Gamma_s$ is non-negligible. Neglecting CP violation in mixing, the relevant replacements for Eq. (15) are [122]

$$\begin{aligned}\Gamma(\overline{B}^0_{s\,\text{phys}} \to f(t)) = \mathscr{N}\tfrac{e^{-t/\tau(B^0_s)}}{4\tau(B^0_s)}\Big[&\cosh(\tfrac{\Delta\Gamma t}{2})+\\ &S_f\sin(\Delta m t) - C_f\cos(\Delta m t) + A_f^{\Delta\Gamma}\sinh(\tfrac{\Delta\Gamma t}{2})\Big],\\ \Gamma(B^0_{s\,\text{phys}} \to f(t)) = \mathscr{N}\tfrac{e^{-t/\tau(B^0_s)}}{4\tau(B^0_s)}\Big[&\cosh(\tfrac{\Delta\Gamma t}{2})-\\ &S_f\sin(\Delta m t) + C_f\cos(\Delta m t) + A_f^{\Delta\Gamma}\sinh(\tfrac{\Delta\Gamma t}{2})\Big].\end{aligned} \tag{20}$$

where $\mathscr{N}$ is a normalisation factor and

$$A_f^{\Delta\Gamma} = -\frac{2\,\text{Re}(\lambda_f)}{1+|\lambda_f|^2}\,. \tag{21}$$

Note that, by definition,

$$\left(S_f\right)^2 + \left(C_f\right)^2 + \left(A_f^{\Delta\Gamma}\right)^2 = 1\,. \tag{22}$$

Also $A_f^{\Delta\Gamma}$ is a CP-conserving parameter, unlike S_f and C_f (since it appears with the same sign in equations for both $\overline{B}^0_s$ and B^0_s states). Nonetheless, it provides sensitivity to $\arg(\lambda_f)$, which means that interesting results can be obtained from untagged time-dependent analyses (*a.k.a.* effective lifetime measurements [123]).

The formalism of Eq. (20) is usually invoked for the determination of the CP violation phase in B^0_s oscillations, $\phi_s = -2\beta_s$, using $B^0_s \to J/\psi\,\phi$ decays. However, in that case things are complicated even further by the fact that the final state, containing two vector particles, is an admixture of CP-even and CP-odd which must be disentangled by angular analysis.[17] Moreover, there is a potential contribution from S-wave K^+K^- pairs within the ϕ mass window used in the analysis. However, all of these features can be turned to the benefit of the analysis, providing better sensitivity and allowing to resolve an ambiguity in the results [125]. A compilation of the latest results is shown in Fig. 9.[18] Although great progress has been made over the last few years, the experimental precision does not yet challenge the theoretical uncertainty, and so further updates are of great interest.

[17] A somewhat more straightforward analysis can be done with the $B^0_s \to J/\psi\, f_0(980)$ decay [124].

[18] Very recent LHCb [130] and ATLAS [131] updates are not included.

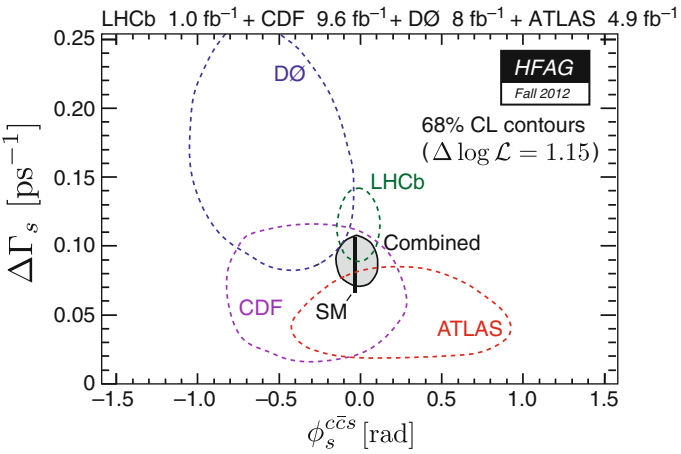

Fig. 9 Compilation of the latest results [30] in the ϕ_s vs. $\Delta\Gamma_s$ plane from LHCb [126], CDF [127], D0 [128] and ATLAS [129]

5.4 *Mixing-Induced CP Violation in Hadronic $b \to s$ Penguin Decay Modes*

As discussed in Sect. 5.1, decay modes mediated by penguin diagrams are potentially sensitive to NP effects, although it is a considerable challenge to disentangle QCD effects. One useful approach is to study mixing-induced CP violation effects in channels that are dominated by the penguin transition, so that little or no tree (or other) contribution is expected. Such channels include $B^0 \to \phi K^0_S$, $B^0 \to \eta' K^0_S$, $B^0 \to K^0_S K^0_S K^0_S$, $B^0_s \to \phi\phi$ and $B^0_s \to K^{*0}\overline{K}^{*0}$.[19] For the B^0 decays, the formalism is the same as given in Eq. (15), and the parameters are expected in the SM to be given, to good approximations, by $C_f \approx 0$, $S_f \approx \sin(2\beta)$ (up to a sign, given by the CP eigenvalue of the final state). These channels have been studied extensively by BaBar and Belle: early results provided hints for discrepancies with the SM predictions, but the significance of the deviation diminished with improved results [132–136]. For the B^0_s decays, the formalism is as given in Eq. (20) (though with modifications due to the vector-vector final states), and the SM expectation is that CP violation effects vanish, to a good approximation, since the very small phase in the $b \to s$ decay cancels that in the B^0_s–$\overline{B}^0_s$ oscillations. First results have been reported by LHCb [137–139], and will reach a very interesting level of sensitivity as more data is accumulated.

[19]The decay $B^0 \to K^0_S\pi^0$ is also of great interest since the tree contribution can be controlled using isospin relations to other $B \to K\pi$ decays.

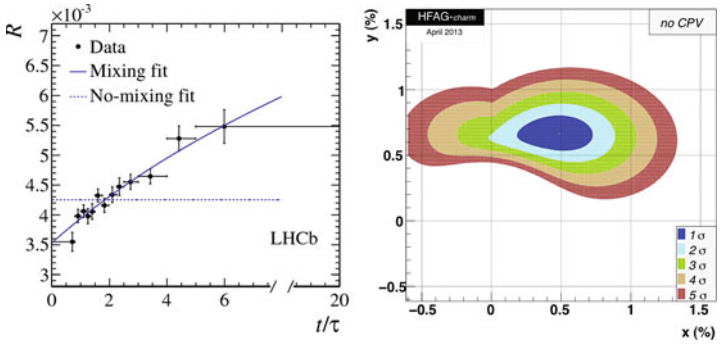

Fig. 10 (*Left*) Decay-time evolution of the ratio, R, of $D^0 \to K^+\pi^-$ to $D^0 \to K^-\pi^+$ yields (*points*) with the projection of the mixing allowed (*solid line*) and no-mixing (*dashed line*) fits overlaid, from Ref. [140]; (*Right*) World average constraints on the x and y parameters in the D^0 system [30]

5.5 Charm Mixing and CP Violation

In the charm system the mixing parameters $x = \Delta m/\Gamma$ and $y = \Delta\Gamma/(2\Gamma)$ are both small, $x, y \ll 1$. Therefore, a Taylor expansion can be performed on the generic expression of Eq. (20) to give

$$\begin{aligned}\Gamma(\overline{D}^0{}_{\text{phys}} \to f(t)) &= \mathcal{N}\frac{e^{-t/\tau(D^0)}}{4\tau(D^0)}\left[1 - C_f + \left(S_f x + A_f^{\Delta\Gamma} y\right)\Gamma t\right],\\ \Gamma(D^0{}_{\text{phys}} \to f(t)) &= \mathcal{N}\frac{e^{-t/\tau(D^0)}}{4\tau(D^0)}\left[1 + C_f - \left(S_f x - A_f^{\Delta\Gamma} y\right)\Gamma t\right].\end{aligned} \tag{23}$$

Hence an untagged analysis of $D^0 \to K^+K^-$ can measure $A_f^{\Delta\Gamma} y$ (also known as y_{CP}), while a tagged analysis can additionally probe $S_f x$. Since the mixing parameters are small, the focus until now has been to establish definitively oscillation effects, but in the coming years the main objective will be to observe or limit CP violation in the charm system, which is expected to be very small in the SM. Note that in case the source of D^0 mesons is either from D^{*+} decays or semileptonic b-hadron decays, the flavour tagging is very effectively achieved from the charge of the associated pion or lepton, respectively. Many other final states can be used to gain additional sensitivity to charm mixing and CP violation parameters, a recent example being the observation of charm mixing at LHCb using $D^0 \to K^+\pi^-$ decays [140]. The result of this analysis, and the world average constraints on the x and y parameters in the D^0 system,[20] are shown in Fig. 10.

[20]Note that in Fig. 10, the definition of the x and y parameters in the charm system is different from that in Sect. 2.2 – in this definition the CP violating phase in D^0 oscillations is assumed to be small, and x can be either positive or negative.

Direct CP violation in the charm system can also be used to test the SM. One interesting recent result has been the measurement of ΔA_{CP}, which is the difference between the direct CP violation parameters of $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays. By measuring the difference, a cancellation of production and detection asymmetries can be exploited, while the physical CP asymmetry may be enhanced.[21] This method was first used by LHCb [141] and then by CDF [142] and Belle [143], all indicating a larger than expected effect. This prompted a great deal of theoretical activity, summarised in Ref. [144], with the conclusion that a SM origin of the CP violation, although unlikely, was not ruled out. Many further studies were proposed to test both SM and NP hypotheses, and these remain of great interest and will be pursued. However, the most recent results by LHCb [145, 146] suggest that the central value is smaller than previous thought, and therefore the SM explanation becomes harder to rule out.

5.6 Photon Polarisation in Radiative B Decays

The $b \to s\gamma$ transition is an archetypal flavour-changing neutral-current (FCNC) transition, and has been considered a sensitive probe for NP since the first measurements of its rate [147, 148]. The latest results for the inclusive branching fraction [30] are consistent with the SM prediction [149]

$$\mathscr{B}(B \to X_s\gamma)^{\rm exp}_{E_\gamma > 1.7\,{\rm GeV}} = (3.43 \pm 0.21 \pm 0.07) \times 10^{-4}\,, \tag{24}$$

$$\mathscr{B}(B \to X_s\gamma)^{\rm th}_{E_\gamma > 1.7\,{\rm GeV}} = (3.15 \pm 0.23) \times 10^{-4}\,. \tag{25}$$

However, additional observables, such as CP and isospin asymmetries provide complementary sensitivity and still have experimental uncertainties much larger than those of the theoretical predictions of their values in the SM.

One particularly interesting observable is the polarisation of the emitted photon in $b \to s\gamma$ decays, since the $V - A$ structure of the SM weak interaction results in a high degree of polarisation, that is not necessarily reproduced in extended models. Until now, the most promising approach to probe the polarisation has been from time-dependent asymmetry measurements of $B^0 \to K^0_S\pi^0\gamma$ [150, 151] but the most recent measurements [152, 153] still have large uncertainties. LHCb can use several different methods to study photon polarisation in $b \to s\gamma$ transitions, such as measuring the effective lifetime in $B^0_s \to \phi\gamma$ decays [154]. Although all such measurements are highly challenging, the observed yields in $B^0_s \to \phi\gamma$ [155] and other related channels such as $B^0 \to K^{*0}e^+e^-$ [156] suggest there are good prospects for significant progress in the coming years.

[21] The CP asymmetries in $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays are expected to be of opposite sign due to U-spin symmetry.

5.7 Angular Observables in $B^0 \to K^{*0}\mu^+\mu^-$ Decays

The $b \to s l^+ l^-$ FCNC transitions offer similar, but complementary, sensitivity to NP as $b \to s\gamma$, but are experimentally more convenient to study, in particular when the lepton pair is muonic, i.e. $l^+l^- = \mu^+\mu^-$. The multi-body final state provides access to a wide range of kinematic observables, several of which have clean theoretical predictions (especially at low values of the dilepton invariant mass squared, q^2). This makes these decays a superb laboratory for NP tests.

The theoretical framework for these (and other) processes is the operator product expansion. This is an effective theory, applicable for b physics, which describes the weak interactions of the SM by integrating out the heavier (W, Z, t) fields. As such it can be considered a modern version of the Fermi theory of beta decay. Conceptually, it can be expressed as

$$\mathscr{L}_{(\text{full EW}\times\text{QCD})} \longrightarrow \mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{QED}\times\text{QCD}}\begin{pmatrix}\text{quarks} \neq \text{t}\\ \text{leptons}\end{pmatrix} + \Sigma_n C_n \mathscr{O}_n\,, \qquad (26)$$

where $\mathscr{O}_n$ represent the local interaction terms, and C_n are coupling constants that are referred to as Wilson coefficients.[22] The Wilson coefficients encode information on the weak scale, and are calculable and known in the SM (at least to leading order). Moreover, they are affected by NP, so comparing the measured values with their expectations provides tests of the SM. A more detailed description of the operator product expansion can be found in, e.g. Ref. [157].

For the purposes of discussing $b \to s l^+ l^-$ decays, the Wilson coefficients of interest are C_7 (which also affects $b \to s\gamma$), C_9 and C_{10}. The differential decay distribution, for the inclusive process, can be written [158]

$$\frac{d^2\Gamma}{dq^2\, d\cos\theta_l} = \frac{3}{8}\left[(1+\cos^2\theta_l)H_T(q^2) + 2\cos\theta_l H_A(q^2) + 2(1-\cos^2\theta_l)H_L(q^2)\right] \qquad (27)$$

where θ_l is the angle between the momentum vectors of the positively charged lepton and the opposite of the decaying b hadron in the dilepton rest frame.[23] The coefficients are given by

$$H_T(q^2) \propto 2q^2\left[\left(C_9 + 2C_7\frac{m_b^2}{q^2}\right)^2 + C_{10}^2\right]$$

[22] As written here the C_n include the Fermi coupling and the CKM matrix elements, but usually these terms are factored out.

[23] The full decay distribution for $B^0 \to K^{*0}\mu^+\mu^-$ and other $B \to V l^+ l^-$ ($V = \rho, \omega, K^*, \phi$) decays includes two other angles: the decay angle of the vector meson (usually denoted θ_V) and the angle between the two decay planes (usually denoted ϕ).

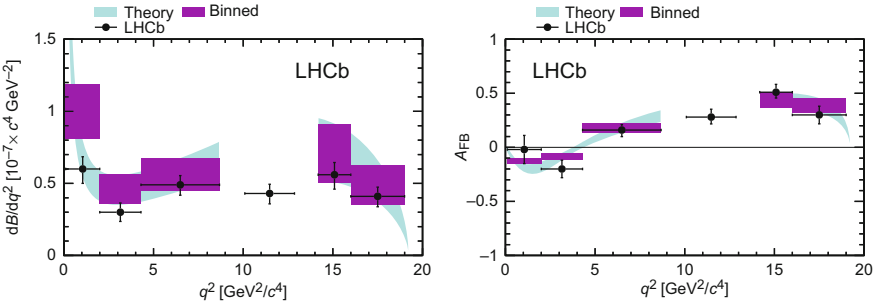

Fig. 11 (*Left*) Differential branching fraction and (*right*) $A_{\rm FB}$ of $B^0 \to K^{*0}\mu^+\mu^-$ decays in bins of q^2 as measured by LHCb [159]

$$H_A(q^2) \propto -4q^2 C_{10}\left(C_9 + 2C_7\frac{m_b^2}{q^2}\right) \qquad (28)$$
$$H_L(q^2) \propto \left[(C_9 + 2C_7)^2 + C_{10}^2\right].$$

Note that the term involving H_A depends linearly on $\cos\theta_l$ and hence gives rise to a q^2-dependent forward backward asymmetry, $A_{\rm FB}$. The shape of $A_{\rm FB}$, in particular the value of q^2 at which it crosses zero, can be predicted with low uncertainty in the SM. The expressions for exclusive processes, such as $B^0 \to K^{*0}\mu^+\mu^-$, are conceptually similar to those of Eqs. (27) and (28), but are more complicated as they also involve hadronic form factors. On the other hand, exclusive channels also provide additional observables that can be studied (such as the longitudinal polarisation of the K^{*0} meson, $F_{\rm L}$), some of which can be precisely predicted in the SM, and are sensitive to NP contributions.

The decay rates and angular distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays have been studied by many experiments, with the most precise results to date, from LHCb [159], shown in Fig. 11. This analysis provides the first measurement of the $A_{\rm FB}$ zero crossing point, $q_0^2 = 4.9 \pm 0.9\,{\rm GeV}^2/c^4$, consistent with the SM prediction. Significant progress, including improved measurements of other NP-sensitive angular observables, can be expected in the coming years.

5.8 *The Very Rare Decay $B_s^0 \to \mu^+\mu^-$*

The "killer app." for flavour physics as a tool to probe for (and potentially discover) NP is the very rare decay $B_s^0 \to \mu^+\mu^-$. The branching fraction is highly suppressed in the SM due to a combination of three factors, none of which are necessarily reproduced in extended models: (i) the absence of tree-level FCNC transitions; (ii) the $V - A$ structure of the weak interaction that results in helicity suppression

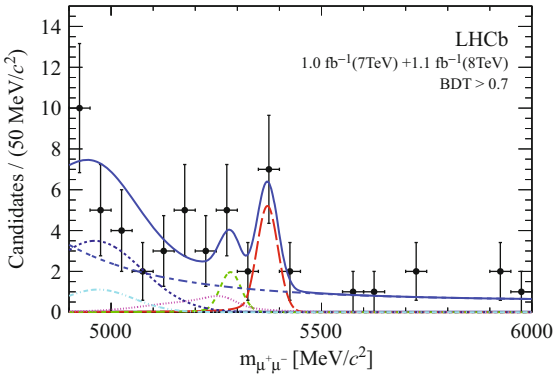

Fig. 12 Invariant mass distribution of selected $B_s^0 \to \mu^+\mu^-$ candidates, with fit result overlaid [167]

of purely leptonic decays of flavoured pseudoscalar mesons; (iii) the hierarchy of the CKM matrix elements. In particular, in the minimally supersymmetric extension of the SM, the presence of a pseudoscalar Higgs particle alleviates the helicity suppression and enhances the branching fraction by a factor proportional to $\tan^6\beta/M_{A_0}^4$, where $\tan\beta$ is the ratio of Higgs' vacuum expectation values, and M_{A_0} is the pseudoscalar Higgs mass. Therefore, in the region of phase-space where $\tan\beta$ is not too small, and M_{A_0} is not too large, the decay rate can be significantly enhanced above its SM expectation [160],[24]

$$\mathscr{B}\left(B_s^0 \to \mu^+\mu^-\right)^{\mathrm{SM}} = (3.2 \pm 0.3) \times 10^{-9}\,. \tag{29}$$

Due to the very clean signature of this decay, it has been studied by essentially all high-energy hadron collider experiments. The crucial components to obtain good sensitivity are high luminosity, a large B production cross-section within the acceptance, and good vertex and mass resolution to reject the background. Although ATLAS [162] and CMS [163] have collected more luminosity, at present the strengths of the LHCb detector have allowed it to obtain the most precise results for this mode. Following a series of increasingly restrictive upper limits [164–166], LHCb has recently obtained the first evidence, with 3.5σ significance, for the decay [167], as shown in Fig. 12. The branching fraction is measured to be

$$\mathscr{B}\left(B_s^0 \to \mu^+\mu^-\right) = \left(3.2\,^{+1.4}_{-1.2}(\mathrm{stat})\,^{+0.5}_{-0.3}(\mathrm{syst})\right) \times 10^{-9}\,, \tag{30}$$

in agreement with the SM prediction.

[24]Note that, due to the non-zero value of the decay width difference in the B_s^0 system, this value needs to be corrected upwards by $\sim$9 % to obtain the experimentally measured (i.e., decay time integrated) quantity [161].

Further updates of this measurement are keenly anticipated, and are likely to appear at regular intervals throughout the lifetime of the LHC. It is worth noting that even in case the $B_s^0 \to \mu^+\mu^-$ branching fraction remains consistent with the SM, the decay provides an additional handle on NP through its effective lifetime [168]. Moreover, it will be important to study also the even more suppressed $B^0 \to \mu^+\mu^-$ decay, since the ratio of the B^0 and B_s^0 branching fractions is a benchmark test of MFV.

6 Future Flavour Physics Experiments

As stressed in the previous sections, the first results from the LHC have already provided dramatic advances in flavour physics, and significant further progress is anticipated in the coming years. However, the instantaneous luminosity of LHCb is limited due to the fact that its subdetectors are read out at 1 MHz. As shown in Fig. 13 (left), increasing the luminosity requires tightening of the hardware trigger thresholds in order not to exceed this limit. This then results in lower efficiencies, especially for decay channels triggered by the calorimeter (i.e., channels without muons in the final state), so that there is no net gain in yield. Therefore, after several years of operation at the optimal instantaneous luminosity at $\sqrt{s} = 13$ or 14 TeV,[25] the time required to double the accumulated statistics becomes excessively long.

As should be clear from the discussions above, it remains of great importance to pursue a wide range of flavour physics measurements and improve their precision to the level of the theoretical uncertainty, and therefore it is of clear interest to get

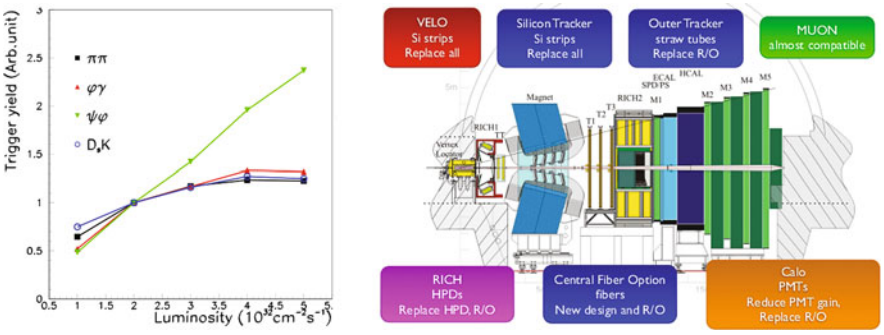

Fig. 13 (*Left*) Scaling of yields with instantaneous in certain decay channels at LHCb [169], showing the limitation caused by the 1 MHz readout. Note that during 2012 LHCb operated at an instantaneous luminosity of $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. (*Right*) Illustration of the key components of the LHCb subdetector upgrades

[25] Note that heavy flavour cross-sections increase with $\sqrt{s}$, so a significant boost in yields is expected when moving to higher energies.

past the 1 MHz readout limitation. The concept of the LHCb upgrade [169, 170] is to read out the full detector at 40 MHz (which corresponds to the maximum bunch crossing rate, with 25 ns spacing) and implement the trigger fully in software. This will allow the experiment to run at higher luminosities, up to 1 or $2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, and will also significantly improve the efficiency for modes currently triggered by calorimeter signals at the hardware level. The accumulated samples in most key modes will increase by around two orders of magnitude compared to what was collected in 2011. Moreover, with a flexible trigger scheme, the capability to search for other signatures of NP will be enhanced, so that the upgraded experiment can be considered a general purpose detector in the forward region. The LHCb upgrade is planned to occur during the second long shutdown of the LHC, in 2018. Since its target luminosity is still below that which can be delivered by the LHC, it does not depend (though it is consistent with) the HL-LHC machine upgrade.

There are several other flavour physics experiments that will be coming online on a similar same timescale. The KEKB accelerator and Belle experiment are being upgraded [171], in order to allow luminosities almost two orders of magnitude larger than has previously been achieved. Compared to the LHCb upgrade, the e^+e^- environment is advantageous for decay modes with missing energy and for inclusive measurements. Some of the key channels for Belle2 are lepton flavour violating decays of τ leptons, mixing-induced CP asymmetries in decays such as $B^0 \to \phi K^0_{\mathrm{S}}$ and $B^0 \to \eta' K^0_{\mathrm{S}}$, and the leptonic decay $B^+ \to \tau^+ \nu$ (which can be considered a counterpart of $B^0_s \to \mu^+\mu^-$, and is sensitive to the exchange of charged Higgs particles) [1, 172].

In addition, the NA62 [173] and K0T0 [174] experiments will search for the $K^+ \to \pi^+ \nu\overline{\nu}$ and $K^0_{\mathrm{L}} \to \pi^0 \nu\overline{\nu}$ decays, respectively. Long considered the "holy grail" of kaon physics these decays are highly suppressed in the SM and have clean theoretical predictions. The new generation of experiments should be able to observe these channels for the first time, if they occur at around the SM rate.

7 Conclusion

Flavour physics continues to present many mysteries, and these demand continued experimental and theoretical investigation. Heavy flavour physics is complementary to other sectors of the global particle physics programme such as the high-p_{T} experiments at the LHC, and neutrino oscillation and low energy precision experiments. The prospects are good for significant progress in the coming few years and, with upgraded experiments planned to come online in the second half of this decade, beyond.

Acknowledgements These lectures were delivered, in variously modified forms, at the Hadron Collider Physics Summer School 2010 in Fermilab, USA, the 2012 Spring School "Bruno Touschek" of the Frascati National Laboratories in Frascati, Italy and the 69th Scottish Universities Summer School in Physics in St. Andrews, UK, 2012. The author is grateful to the organisers

of these meetings for invitations and support, and to the participants for stimulating discussions and questions. This work is supported by the Science and Technology Facilities Council (United Kingdom), CERN, and by the European Research Council under FP7.

References

1. T.E. Browder, T. Gershon, D. Pirjol, A. Soni, J. Zupan, Rev. Mod. Phys. **81**, 1887 (2009)
2. N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963)
3. M. Kobayashi, T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973)
4. I.I. Bigi, AIP Conf. Proc. **814**, 230 (2006)
5. S.L. Glashow, J. Iliopoulos, L. Maiani, Phys. Rev. D **2**, 1285 (1970)
6. J. Adam et al., MEG Collaboration, Phys. Rev. Lett. **110**, 201801 (2013)
7. W.J. Marciano, T. Mori, J.M. Roney, Ann. Rev. Nucl. Part. Sci. **58**, 315 (2008)
8. L.D. Landau, Nucl. Phys. **3**, 127 (1957)
9. T.D. Lee, C.-N. Yang, Phys. Rev. **104**, 254 (1956)
10. C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes, R.P. Hudson, Phys. Rev. **105**, 1413 (1957)
11. G. Luders, Ann. Phys. **2**, 1 (1957) [Ann. Phys. **281**, 1004 (2000)]
12. A. Angelopoulos et al., CPLEAR Collaboration, Phys. Lett. B **444**, 43 (1998)
13. J.P. Lees et al., BaBar Collaboration, Phys. Rev. Lett. **109**, 211801 (2012)
14. L.-L. Chau, W.-Y. Keung, Phys. Rev. Lett. **53**, 1802 (1984)
15. L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983)
16. J.H. Christenson, J.W. Cronin, V.L. Fitch, R. Turlay, Phys. Rev. Lett. **13**, 138 (1964)
17. A.D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967) [JETP Lett. **5**, 24 (1967)]; [Sov. Phys. Usp. **34**, 392 (1991)]; [Usp. Fiz. Nauk **161**, 61 (1991)]
18. P. Dirac, Theory of electrons and positrons, 1933, available from http://www.nobelprize.org/
19. C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985)
20. S. Davidson, E. Nardi, Y. Nir, Phys. Rep. **466**, 105 (2008)
21. B. Pontecorvo, Sov. Phys. JETP **6**, 429 (1957) [Zh. Eksp. Teor. Fiz. **33**, 549 (1957)]
22. Z. Maki, M. Nakagawa, S. Sakata, Prog. Theor. Phys. **28**, 870 (1962)
23. F.P. An et al., DAYA-BAY Collaboration, Phys. Rev. Lett. **108**, 171803 (2012)
24. J.K. Ahn et al., RENO Collaboration, Phys. Rev. Lett. **108**, 191802 (2012)
25. See the review by D. Kirkby and Y. Nir in Ref. [28]
26. See the review by O. Schneider in Ref. [28]
27. U. Nierste, (2009), arXiv:0904.1869 [hep-ph]
28. J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012)
29. See the review by L. Wolfenstein, C.-J. Lin and T.G. Trippe in Ref. [28]
30. Y. Amhis et al., Heavy Flavor Averaging Group Collaboration, (2012), arXiv:1207.1158 [hep-ex]; updated results and plots available at: http://www.slac.stanford.edu/xorg/hfag/
31. G. Isidori, Y. Nir, G. Perez, Ann. Rev. Nucl. Part. Sci. **60**, 355 (2010)
32. A. Lenz, U. Nierste, J. Charles, S. Descotes-Genon, H. Lacker, S. Monteil, V. Niess, S. T'Jampens, Phys. Rev. D **86**, 033008 (2012)
33. G. D'Ambrosio, G.F. Giudice, G. Isidori, A. Strumia, Nucl. Phys. B **645**, 155 (2002)
34. V.M. Abazov et al., D0 Collaboration, Phys. Rev. D **84**, 052007 (2011)
35. A. Lenz, U. Nierste, (2011), arXiv:1102.4274 [hep-ph]
36. See the review by E. Blucher and W. Marciano in Ref. [28]
37. J. Charles et al., CKMfitter Group Collaboration, Eur. Phys. J. C **41**, 1 (2005), [hep-ph/0406184]; updated results and plots available at: http://ckmfitter.in2p3.fr
38. M. Bona et al., UTfit Collaboration, JHEP **0507**, 028 (2005);updated results and plots available at: http://www.utfit.org/UTfit/
39. A.J. Buras, M.E. Lautenbacher, G. Ostermaier, Phys. Rev. D **50**, 3433 (1994)
40. A.B. Carter, A.I. Sanda, Phys. Rev. D **23**, 1567 (1981)

41. I.I.Y. Bigi, A.I. Sanda, Nucl. Phys. B **193**, 85 (1981)
42. G.C. Branco, L. Lavoura, J.P. Silva, Int. Ser. Monogr. Phys. **103**, 1 (1999)
43. I.I.Y. Bigi, A.I. Sanda, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **9**, 1 (2000)
44. B. Aubert et al., BaBar Collaboration, Nucl. Instrum. Method A **479**, 1 (2002)
45. B. Aubert et al., BaBar Collaboration, Nucl. Instrum. Method A **729**, 615 (2013)
46. A. Abashian et al., Nucl. Instrum. Method A **479**, 117 (2002)
47. B. Aubert et al., BaBar Collaboration, Phys. Rev. Lett. **87**, 091801 (2001)
48. K. Abe et al., Belle Collaboration, Phys. Rev. Lett. **87**, 091802 (2001)
49. B. Aubert et al., BaBar Collaboration, Phys. Rev. D **79**, 072009 (2009)
50. I. Adachi et al., Belle Collaboration, Phys. Rev. Lett. **108**, 171802 (2012)
51. M. Gronau, D. London, Phys. Rev. Lett. **65**, 3381 (1990)
52. B. Aubert et al., Babar Collaboration, Phys. Rev. D **76**, 052007 (2007)
53. A. Somov et al., Belle Collaboration, Phys. Rev. D **76**, 011104 (2007)
54. K. Abe et al., Belle Collaboration, Phys. Rev. D **71**, 072003 (2005) [Erratum-ibid. D **71**, 079903 (2005)]
55. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **719**, 318 (2013)
56. A. Abulencia et al., CDF Collaboration, Phys. Rev. Lett. **97**, 242003 (2006)
57. R. Aaij et al., LHCb Collaboration, New J. Phys. **15**, 053021 (2013)
58. V.G. Luth, Ann. Rev. Nucl. Part. Sci. **61**, 119 (2011)
59. P. del Amo Sanchez et al., BaBar Collaboration, Phys. Rev. D **83**, 052011 (2011)
60. P. del Amo Sanchez et al., BaBar Collaboration, Phys. Rev. D **83**, 032007 (2011)
61. H. Ha et al., Belle Collaboration, Phys. Rev. D **83**, 071101 (2011)
62. V. Gibson, Lectures presented at the Fourth CERN-Fermilab Hadron Collider Physics Summer School, CERN, Geneva, 2009
63. A.A. Alves Jr. et al., LHCb Collaboration, JINST **3**, S08005 (2008)
64. R. Aaij et al., JINST **8**, P04022 (2013)
65. L. Ristori, G. Punzi, Ann. Rev. Nucl. Part. Sci. **60**, 595 (2010)
66. D. Acosta et al., CDF Collaboration, Phys. Rev. D **71**, 032001 (2005)
67. G. Aad et al., ATLAS Collaboration, Nucl. Phys. B **850**, 387 (2011)
68. S. Chatrchyan et al., CMS Collaboration, JHEP **1202**, 011 (2012)
69. R. Aaij et al., LHCb Collaboration, Eur. Phys. J. C **71**, 1645 (2011)
70. R. Aaij et al., LHCb Collaboration, JHEP **1302**, 041 (2013)
71. R. Aaij et al., LHCb Collaboration, JHEP **1306**, 064 (2013)
72. T. Aaltonen et al., CDF Collaboration, Phys. Rev. D **77**, 072004 (2008)
73. T. Aaltonen et al., CDF Collaboration, Phys. Rev. D **79**, 092003 (2009)
74. S. Abachi et al., D0 Collaboration, Phys. Rev. Lett. **74**, 3548 (1995)
75. G. Aad et al., ATLAS Collaboration, Nucl. Phys. B **864**, 341 (2012)
76. S. Chatrchyan et al., CMS Collaboration, JHEP **1206**, 110 (2012)
77. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **694**, 209 (2010)
78. D. Acosta et al., CDF Collaboration, Phys. Rev. Lett. **91**, 241804 (2003)
79. R. Aaij et al., LHCb Collaboration, Nucl. Phys. B **871**, 1 (2013)
80. R. Aaij et al., LHCb Collaboration, Phys. Rev. D **85**, 032008 (2012)
81. S.K. Choi et al., Belle Collaboration, Phys. Rev. Lett. **91**, 262001 (2003)
82. B. Aubert et al., BaBar Collaboration, Phys. Rev. Lett. **90**, 242001 (2003)
83. D. Besson et al., CLEO Collaboration, Phys. Rev. D **68**, 032002 (2003) [Erratum-ibid. D **75**, 119908 (2007)]
84. G. Aad et al., ATLAS Collaboration, Phys. Rev. Lett. **108**, 152001 (2012)
85. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. Lett. **108**, 252002 (2012)
86. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **109**, 172003 (2012)
87. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **110**, 222001 (2013)
88. S.K. Choi et al., Belle Collaboration, Phys. Rev. Lett. **100**, 142001 (2008)
89. R. Mizuk et al., Belle Collaboration, Phys. Rev. D **80**, 031104 (2009)
90. R. Mizuk et al., Belle Collaboration, Phys. Rev. D **78**, 072004 (2008)
91. B. Aubert et al., BaBar Collaboration, Phys. Rev. D **79**, 112001 (2009)

92. J.P. Lees et al., BaBar Collaboration, Phys. Rev. D **85**, 052003 (2012)
93. M. Ablikim et al., BESIII Collaboration, Phys. Rev. Lett. **110**, 252001 (2013)
94. Z.Q. Liu et al., Belle Collaboration, Phys. Rev. Lett. **110**, 252002 (2013)
95. A. Bondar et al., Belle Collaboration, Phys. Rev. Lett. **108**, 122001 (2012)
96. I. Adachi et al., Belle Collaboration, (2012). arXiv:1209.6450 [hep-ex]
97. D.C. Hom et al., Phys. Rev. Lett. **36**, 1236 (1976)
98. B. Aubert et al., BaBar Collaboration, Phys. Rev. Lett. **101**, 071801 (2008) [Erratum-ibid. **102**, 029901 (2009)]
99. S.W. Herb et al., Phys. Rev. Lett. **39**, 252 (1977)
100. J.P. Lees, BaBar Collaboration, Phys. Rev. D **87**(5), 052009 (2013)
101. Y.-T. Duh, et al., Belle Collaboration, Phys. Rev. D **87**, 031103 (2013)
102. R. Aaij et al., LHCb Collaboration, JHEP **1301**, 111 (2013)
103. CDF collaboration, Direct CP Violating Asymmetries in Charmless Decays of Strange Bottom Mesons and Bottom Baryons with 9.3 fb^{-1}. Public Note 10726 2012
104. T. Aaltonen et al., CDF Collaboration, Phys. Rev. Lett. **108**, 211803 (2012)
105. R. Aaij et al., LHCb Collaboration, JHEP **10**, 183 (2013)
106. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **110**, 221601 (2013)
107. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **111**, 101801 (2013)
108. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **112**, 011801 (2014)
109. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **111**, 101801 (2013)
110. R. Aleksan, I. Dunietz, B. Kayser, Z. Phys. C **54**, 653 (1992)
111. T. Gershon, Phys. Rev. D **79**, 051301 (2009)
112. T. Gershon, M. Williams, Phys. Rev. D **80**, 092002 (2009)
113. R. Aaij et al., LHCb Collaboration, LHCb-CONF-2012-029, 2012
114. R. Aaij et al., LHCb Collaboration, Phys. Rev. D **87**, 112009 (2013)
115. J. Brod, J. Zupan, JHEP **01**, 051 (2014)
116. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **712**, 203 (2012) [Erratum-ibid. B **713**, 351 (2012)]
117. J.P. Lees et al., BaBar Collaboration, Phys. Rev. D **87**, 052015 (2013)
118. K. Trabelsi, Belle Collaboration, (2013), arXiv:1301.2033 [hep-ex]
119. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **726**, 151 (2013)
120. R. Aaij et al., LHCb Collaboration, LHCb-CONF-2013-004, 2013
121. R. Aaij et al., LHCb Collaboration, LHCb-CONF-2013-006, 2013
122. I. Dunietz, R. Fleischer, U. Nierste, Phys. Rev. D **63**, 114015 (2001)
123. R. Fleischer, R. Knegjens, G. Ricciardi, Eur. Phys. J. C **71**, 1798 (2011)
124. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **713**, 378 (2012)
125. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **108**, 241801 (2012)
126. R. Aaij et al., LHCb Collaboration, Phys. Rev. **D87**, 112010 (2013)
127. T. Aaltonen et al., CDF Collaboration, Phys. Rev. Lett. **109**, 171802 (2012)
128. V.M. Abazov et al., D0 Collaboration, Phys. Rev. D **85**, 032006 (2012)
129. G. Aad et al., ATLAS Collaboration, JHEP **1212**, 072 (2012)
130. R. Aaij et al., LHCb Collaboration, Phys. Rev. D **87**, 112010 (2013)
131. G. Aad et al., ATLAS Collaboration, ATLAS-CONF-2013-039, 2013
132. J.P. Lees et al., BaBar Collaboration, Phys. Rev. D **85**, 112010 (2012)
133. Y. Nakahama et al., Belle Collaboration, Phys. Rev. D **82**, 073011 (2010)
134. B. Aubert et al., BaBar Collaboration, Phys. Rev. D **79**, 052003 (2009)
135. K.-F. Chen et al., Belle Collaboration, Phys. Rev. Lett. **98**, 031802 (2007)
136. J.P. Lees et al., BaBar Collaboration, Phys. Rev. D **85**, 054023 (2012)
137. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **709**, 50 (2012)
138. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **713**, 369 (2012)
139. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **110**, 241802 (2013)
140. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **110**, 101802 (2013)
141. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **108**, 111602 (2012)
142. T. Aaltonen et al., CDF Collaboration, Phys. Rev. Lett. **109**, 111801 (2012)

143. B.R. Ko, Belle Collaboration, PoS ICHEP **2012**, 353 (2013)
144. R. Aaij et al., LHCb Collaboration, Eur. Phys. J. C **73**, 2373 (2013)
145. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **723**, 33 (2013)
146. R. Aaij et al., LHCb Collaboration, LHCb-CONF-2013-003, 2013
147. R. Ammar et al., CLEO Collaboration, Phys. Rev. Lett. **71**, 674 (1993)
148. M.S. Alam et al., CLEO Collaboration, Phys. Rev. Lett. **74**, 2885 (1995)
149. M. Misiak et al., Phys. Rev. Lett. **98**, 022002 (2007)
150. D. Atwood, M. Gronau, A. Soni, Phys. Rev. Lett. **79**, 185 (1997)
151. D. Atwood, T. Gershon, M. Hazumi, A. Soni, Phys. Rev. D **71**, 076003 (2005)
152. B. Aubert et al., BaBar Collaboration, Phys. Rev. D **78**, 071102 (2008)
153. Y. Ushiroda et al., Belle Collaboration, Phys. Rev. D **74**, 111104 (2006)
154. F. Muheim, Y. Xie, R. Zwicky, Phys. Lett. B **664**, 174 (2008)
155. R. Aaij et al., LHCb Collaboration, Nucl. Phys. B **867**, 1 (2013)
156. R. Aaij et al., LHCb Collaboration, JHEP **1305**, 159 (2013)
157. A.J. Buras, (2005), hep-ph/0505175
158. K.S.M. Lee, Z. Ligeti, I.W. Stewart, F.J. Tackmann, Phys. Rev. D **75**, 034016 (2007)
159. R. Aaij et al., LHCb Collaboration, JHEP **1308**, 131 (2013)
160. A.J. Buras, J. Girrbach, D. Guadagnoli, G. Isidori, Eur. Phys. J. C **72**, 2172 (2012)
161. K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino, N. Tuning, Phys. Rev. Lett. **109**, 041801 (2012)
162. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **713**, 387 (2012)
163. S. Chatrchyan et al., CMS Collaboration, JHEP **1204**, 033 (2012)
164. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **699**, 330 (2011)
165. R. Aaij et al., LHCb Collaboration, Phys. Lett. B **708**, 55 (2012)
166. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **108**, 231801 (2012)
167. R. Aaij et al., LHCb Collaboration, Phys. Rev. Lett. **110**, 021801 (2013)
168. A.J. Buras, R. Fleischer, J. Girrbach, R. Knegjens, JHEP **1307**, 77 (2013)
169. A. Schopper, R. Lindner et al., LHCb collaboration, Framework TDR for the LHCb upgrade: technical design report, CERN-LHCC-2012-007, 2012
170. S. Stone, R. Lindner et al., LHCb collaboration, Letter of intent for the LHCb upgrade, CERN-LHCC-2011-001, 2011
171. T. Abe et al., Belle-II Collaboration, (2010), arXiv:1011.0352 [physics.ins-det]
172. T. Aushev et al., (2010), arXiv:1002.5012 [hep-ex]
173. F. Hahn et al., NA62: technical design document, NA62-10-07, 2010
174. T. Yamanaka, KOTO Collaboration, PTEP **2012**, 02B006 (2012)

Searches for New Physics at the Large Hadron Collider

Jeffrey D. Richman

Abstract The experiments currently underway at the Large Hadron Collider are exploring the physics of the TeV energy scale, which may hold the answers to some of the most profound questions in particle physics. These lectures describe the status of searches for new physics beyond the standard model, focusing on supersymmetry, but addressing other aspects of this enormously broad physics program as well. Such topics as extra spatial dimensions, new gauge bosons, and microscopic black holes are included in the category known as exotica; some of these possibilities are also considered here. The methodologies and challenges associated with searches for new physics are discussed, followed by a survey of some of the basic phenomenology and the experimental results. This pedagogical review is intended for graduate students and postdocs who are working on this critical part of the LHC research program.

1 Introduction

I am delighted to be here at St. Andrews University, a distinguished 600-year-old institution, to present lectures on searches for new physics at the Large Hadron Collider (LHC). The current period is one of intensive effort to explore the physics of the TeV energy scale, which may allow us to address some of the most fundamental mysteries of nature. Figure 1 shows a conception of the particle physics landscape, both known and speculative, by Sergio Cittolin, a fellow member of CMS. We are all privileged as scientists to be able to use one of the most extraordinary scientific instruments of all time – the LHC – to explore the unknown territory of the TeV scale.

In these lectures, I adopt an unashamedly pedagogical approach. My goal is to explain as many simple things as possible, to focus on topics that I find interesting and fun, and not to worry about being comprehensive and balanced, as one would be in a review talk at a conference. I have tried to avoid covering results and ideas that are explained by other speakers at this school. Peter Maettig has done an admirable

J.D. Richman (✉)
Department of Physics, University of California, Santa Barbara, CA 93106, USA
e-mail: richman@hep.ucsb.edu

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_7

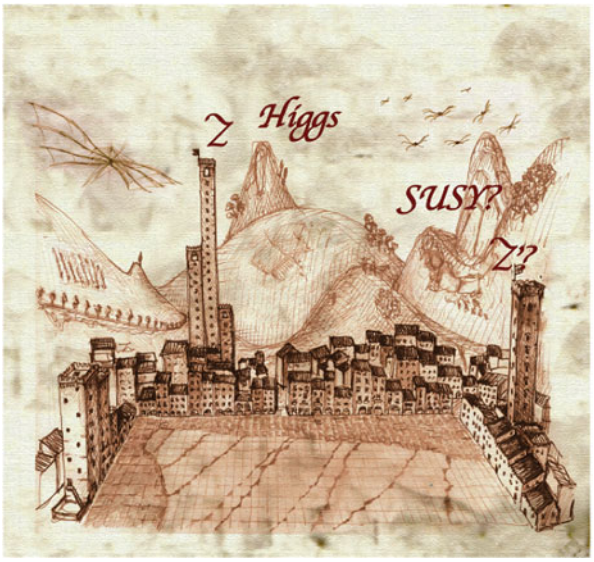

Fig. 1 A view of the LHC particle physics landscape by Sergio Cittolin, in the style of Leonardo da Vinci (Figure used with permission)

job of describing many important new physics searches in the context of his lectures on standard model (SM) results from the LHC experiments. Giacomo Cacciapaglia and Sven Heinemeyer have presented beautiful lectures on theoretical aspects of new physics. Bill Murray has summarized the amazing experimental results from the Higgs searches. For an excellent discussion of statistical issues, the reader is referred to the lectures by Glen Cowan. My own presentation is shaped by my involvement in supersymmetry searches in the CMS experiment, but I have tried to use examples from ATLAS as well.

2 Key Problems and Puzzles at the Electroweak Scale

This is a special year for physics and for the Scottish Universities Summer School in Physics. The discovery of a new particle [1, 2] with mass $m \approx 125\,\text{GeV}$ and properties consistent with those of a Higgs boson is an historic achievement. We are honored to have Peter Higgs with us here to speak about the deep insights that he, Robert Brout, Francois Englert, and other theorists developed some 50 years ago [3–6]. These ideas have provided invaluable guidance to our field, and they have helped us to develop powerful experimental tools and methods needed for this remarkable discovery, especially the ATLAS [7] and CMS [8] detectors.

Given that the new particle decays to $\gamma\gamma$ and ZZ, it must be a boson. Furthermore, a spin-1 particle cannot decay into two photons [9], so the new particle cannot be another massive, spin-1 gauge boson, like the Z. Assuming

that the Higgs hypothesis is confirmed by ongoing measurements, the particle will be the first fundamental scalar particle observed in nature (there are, of course, mesons with the quantum numbers $J^{PC} = 0^{++}$ [10]). We are thus on the verge of confirming the mechanism of electroweak symmetry breaking, in which the properties of the vacuum play a crucial role in explaining how massive gauge bosons can be accommodated in a gauge theory without destroying gauge invariance. And in the unlikely scenario in which the new particle turns out not to be a Higgs boson, we will be in a state of complete confusion, which will be even more interesting!

The observation of this new Higgs-like particle suggests strongly that we are on the right track conceptually in particle physics. We have found a new puzzle piece, and it appears to fit perfectly! But while it may turn out that the SM is nominally complete, the discovery certainly does not come close to resolving all of the many profound mysteries of our field.

Although the Higgs sector helps us to understand the origin of fundamental particle masses in a gauge theory, the low mass of the new particle itself presents a puzzle, which has been anticipated for some time. This is the gauge hierarchy problem [11], and it is inextricably bound up with the spin-0 nature of the Higgs boson. The squares of the bare masses (m^0) of fundamental scalar particles generically receive radiative corrections from quantum loop effects that depend quadratically on the cutoff scale Λ for momenta in the loop. In the case of the Higgs boson (h),

$$m_h^2 = (m_h^0)^2 + \Delta m_h^2 \tag{1}$$

and in the SM, the one-loop corrections are [12]

$$\Delta m_h^2 \simeq -\frac{3\Lambda^2}{8\pi^2 v^2}(4m_t^2 - 2M_W^2 - M_Z^2 - m_h^2), \tag{2}$$

where v is the Higgs vacuum expectation value. The contribution from the top-quark loop is dominant for light Higgs masses, and, barring a fine tuning of parameters to arrange cancellation of the loop effects, the Higgs mass is pulled to the scale Λ. If Λ corresponds to a high mass scale such as the reduced Planck scale $M_{\rm Pl} = \sqrt{\hbar c/(8\pi G_N)} \simeq 2.4 \times 10^{18}$ GeV, the degree of fine tuning required is severe, around 30 orders of magnitude. While such a cancellation is not excluded in principle, we are not aware of any physical reason for it to occur, and it seems highly unlikely that such a precise cancellation would occur by accident. This prediction is therefore regarded (by at least by some people) as unnatural. The criterion of *naturalness* is not straightforward to define, and various definitions have been given in the literature [13, 14]. Roughly speaking, the predictions of a natural theory should be stable with respect to small variations in its parameters. As discussed in the lectures by Giacomo Cacciapaglia in these proceedings, there are several possible avenues for stabilizing the mass hierarchy, including supersymmetry, extra dimensions, and technicolour. The discovery of a Higgs-like particle has actually given us more reason to search for new physics beyond the SM!

A perhaps more empirically based reason to search for new physics at the LHC is the compelling evidence for dark matter [15–17], which is known from astrophysical observations to dominate the matter density of the universe. The effects of dark matter are observed in several types of phenomena, including galactic rotation curves, anisotropies in the cosmic microwave background radiation, and microlensing observations. We should be humbled by the fact that, in spite of our excellent understanding of "ordinary" atomic matter (and its relatives in the second and third generations), the majority of the matter of the universe cannot be explained by the SM particle spectrum. To unravel this mystery, we need to detect cosmic dark matter directly, and we need to produce and study dark matter in detail. Ideally, the information from two lines of investigation—special low-background experiments and accelerator-based experiments—would then be combined, giving a full understanding of the physics of dark matter.

Supersymmetry (SUSY), which relates fermions and bosons, is a framework that, in a large range of scenarios, provides not only a solution to the gauge hierarchy problem, but also a dark-matter candidate. SUSY extends the Poincaré group of Lorentz boosts, rotations, and translations in a radical way, mapping bosonic and fermion degrees of freedom onto each other [11,18–23]. The quantum corrections to the Higgs-boson mass are greatly reduced by the presence of amplitudes associated with loop diagrams containing virtual SUSY particles, substantially cancelling the corresponding SM contributions [13]. Another attractive feature of SUSY is that it can lead to convergence of the running gauge coupling constants at high energy, which would be an indication that the physical laws of nature can be described (at high energies) by a unified gauge theory with a single gauge group and a single gauge coupling constant. If the SUSY partners exist at the weak scale, unification of the coupling constants can take place around $M_{\mathrm{GUT}} \sim 2 \times 10^{16}\,\mathrm{GeV}$. For all of these reasons, SUSY has acquired a somewhat special status as an extension to the SM. My own emotional state oscillates between awe at its fundamental beauty and deep implications on the one hand and, on the other, dismay at its complexity, particularly with regard to the breaking of supersymmetry. So many scenarios, so many parameters! But in the end, the only thing that matters is whether the theory describes the real world, not whether we think it is beautiful.

The range of new physics possibilities accessible at the LHC extends far beyond SUSY. The term *exotica*, which is used in both ATLAS and CMS as well as in the Tevatron experiments, encompasses a vast range of new particles and phenomena. These include resonances, such as heavy gauge bosons (W', Z'); compositeness of SM particles (substructure); 4th generation particles; leptoquarks (particles with both lepton and baryon quantum numbers); various scenarios leading to long-lived particles (including SUSY); microscopic black holes (motivated by ideas about TeV-scale gravity); heavy neutrinos; tests of triple gauge couplings; and contact interactions (resulting from the exchange of very heavy particles). It is important to recognize that SUSY is not the only idea for addressing the gauge hierarchy problem. For example, Randall-Sundrum warped-extra-dimension models [24] and models with large extra dimensions [25] provide intriguing alternative perspectives. Some of these exotica searches, including the extremely important possibility of

Table 1 Broad categories for supersymmetry searches. The list is far from comprehensive, and many of the categories overlap

Supersymmetry
R-parity conserving (E_T^{miss}-based searches)
R-parity violating (searches without the E_T^{miss} signature)
Inclusive searches for topological signatures (e.g., for MSUGRA/cMSSM)
Searches for signature for gauge-mediated SUSY breaking
Searches for signatures with γ, Z
Searches motivated by naturalness considerations (light $\tilde{t}$, $\tilde{b}$, $\tilde{g}$, $\tilde{\chi}^{\pm}$, $\tilde{\chi}^{0}$)
Strong production of SUSY
Electroweak production of SUSY
Monojet events and connection to direct dark matter searches
Long-lived SUSY particles, e.g., long-lived gluinos, R-hadrons
Split SUSY
Stealth SUSY

Table 2 A partial list of the main categories for exotica searches

Exotica
Large extra dimensions
Universal extra dimensions
Randall-Sundrum models
Hidden valley models
Microscopic black holes
Contact interactions
New heavy gauge bosons
Leptoquarks
4th generation quarks and leptons
Excited quarks and leptons
Technihadrons
Heavy Majorana neutrinos
Heavy right-handed W bosons
Long-lived particles

new heavy gauge bosons, arise naturally in the context of detailed or even precision studies of SM processes and are discussed in the lectures by Peter Maettig.

Tables 1 and 2 list some of the main categories for SUSY and exotica searches. While these and other motivations for new physics searches are intriguing and even compelling, I would like to advocate that we not lose sight of another more basic perspective. This is simply that the TeV scale is, on empirical grounds, a critical energy scale of nature, and it may provide information that allows us to access physical laws operating at much higher mass scales. Looking back, it required several decades to explore and understand the physics accessible at the GeV scale. That scale yielded far more physics than anyone could have possibly imagined. In fact, the LHCb experiment is still pursuing many important questions in B and B_s

meson physics, some of which have important implications for physics at higher mass scales. The TeV scale could require substantially more time and effort to understand than the GeV scale, and the LHC may not be able to provide all the answers. But it is our responsibility to exploit the full potential of the LHC as we explore this new territory.

These lectures consist of four main parts: methodological challenges and problems in searches for new physics (Sect. 4), characteristics of SM backgrounds (Sect. 5), searches for supersymmetry (Sect. 6), and searches for exotica, focusing on searches with unusual features and methods (Sect. 7).

3 References and Resources

Both ATLAS and CMS maintain web pages that enable one to quickly obtain an overview of the search results currently available. The starting points for obtaining ATLAS and CMS physics results are

- https://twiki.cern.ch/twiki/bin/view/AtlasPublic
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

Results related to supersymmetry are linked to the following web pages:

- https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS

Results related to exotica searches are linked to the following web pages:

- https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsEXO

There are also numerous review articles and books that can be extremely helpful to someone facing the daunting task of learning this physics. A time-honored resource is *A Supersymmetry Primer*, by S. P. Martin [23]. Several detailed books on supersymmetry with extensive discussions of phenomenology are also now available. *Supersymmetry in Particle Physics*, by I. Aitchison [26], is particularly helpful to beginners. Other books, such as *Weak Scale Supersymmetry*, by H. Baer and X. Tata [27], *Supersymmetry: Theory, Experiment, and Cosmology*, by P. Binétruy [12], and *Theory and Phenomenology of Sparticles*, by M. Drees, R. H. Godbole, and P. Roy [28], are more advanced and comprehensive. A standard text that has been used for many years is *Supersymmetry and Supergravity* by J. Wess and J. Bagger [29]. The *Review of Particle Properties* contains two reviews of supersymmetry, one theoretical (H. Haber) [11] and one experimental (Buchmuller and de Jong) [30], each of which provides a wealth of information and references. A valuable resource for understanding the ATLAS and CMS detectors is *At the Leading Edge: the ATLAS and CMS LHC Experiments*, edited by Dan Green [31].

4 Challenges in Searches for New Physics

There are many challenges in searching for new physics. The first part of this section, *Bumps in the Road* (Sect. 4.1), presents examples of searches that have run into difficulties, leading to conclusions that were not confirmed by later studies. In the second part, *Lessons Learned: Common Problems in Searches for New Physics* (Sect. 4.2), we consider what lessons can be learned from these struggles.

Here are some questions to think about:

- How well do you understand the detector systems and software that affect your measurement?
- Are there aspects of the analysis that are not validated by studies of control samples in the data? What is the weakest element of the analysis? Can this element be strengthened?
- If you observe an excess in your search, will you trust the systematic uncertainties and the significance or will you want to rethink them? What are the systematic uncertainties fundamentally based on?

And here are some provocative assertions to consider:

- The foundation of any search is a detailed understanding of the SM backgrounds.
- In a well-designed and executed analysis, one obtains a coherent physical picture of the event sample, including both the signal region and the surrounding neighborhood in the sample. This physical picture gives credibility to the results.
- In general, systematic uncertainties do not have a well-defined probability content. If they are comparable to (or larger than) the statistical uncertainties, the meaning of the total uncertainty becomes questionable.

4.1 Bumps in the Road

Searches for new particles or for new physical processes present both great opportunities and challenges. Historically, a significant number of searches in our field have encountered serious problems, and it is instructive to consider some of these and to see what lessons we can learn. Below we review examples of measurements that obtained conclusions that were later found to be incorrect. I will not, however, try to give a detailed explanation of what happened in each case –it is not always straightforward to obtain a clear picture from the published literature. I encourage you to read the papers and to develop your own ideas.

In 1984, the Crystal Ball experiment announced a preliminary result [32] (which was not published in a journal), "Evidence for a Narrow Massive State in the Radiative Decays of the Upsilon." The full process was $e^+e^- \to \Upsilon(1S) \to \gamma + \zeta(8.3)$, where $\zeta(8.3)$ was the name tentatively assigned to a new particle whose mass (8.3 GeV) was inferred from the location of the monochromatic peak in the energy spectrum of the recoiling photon. The $\Upsilon(1S)$ is a $b\bar{b}$ bound state with $L = 0$,

$S = 1$, and $J = L + S = 1$ (in the sense of addition of angular momenta). That is, there is no relative orbital angular momentum between the b and $\bar{b}$ quarks ($L = 0$), but the two quark spins (each 1/2) are coupled in a symmetric state to a spin of $S = 1$, which is also the total spin J of the meson (since $L = 0$). The key element of the detector was an array of NaI crystals, which provided excellent photon-energy resolution. The experiment did not have a magnet, so the information on charged tracks was limited, a weakness of the measurement. Compared with the events typically observed at the LHC, those in this study had a very low particle multiplicity: the initial state consisted of a single particle, the $\Upsilon(1S)$.

The physics of $b\bar{b}$ and $c\bar{c}$ states may seem less familiar today, but the Crystal Ball experiment had a long and distinguished history at SLAC of mapping out the states in the charmonimum system by studying the spectroscopy of the radiative transitions between $c\bar{c}$ states of different quantum numbers. The detector was subsequently moved (in a U.S. Air Force transport plane) to Hamburg, where it was installed in the DORIS ring at DESY. This ring operated at a higher energy to study the Υ system as well as B mesons. The radiative decays of both $c\bar{c}$ and $b\bar{b}$ states have provided extensive information on these "onia" particles.

The Crystal Ball evidence for a state at mass of 8.3 GeV consisted of two separate photon energy spectra, each with a peak just above 1.0 GeV, as shown in Fig. 2. The two samples were separated on the basis of the characteristics of the hadronic recoil system. In the $\zeta(8.3) \rightarrow$ *multi-hadron channel*, the statistical significance was 4.2σ, while in the $\zeta(8.3) \rightarrow$ *two jets channel*, the significance was 3.3σ.

The March 1985 issue of *Physics Today* [33] contains an article entitled *Zeta revisited: Have we really seen the Higgs?* The article begins

> Much excitement was generated last summer at the XXIII International Conference on High Energy Physics in Leipzig by the Crystal Ball collaboration's report of evidence for a curious new particle, the 8.32-GeV "zeta" boson, that might well have been the long-sought-after Higgs particle.

Let's consider some of the strengths and weaknesses of the signature. A strength is certainly the narrow peak in a kinematic quantity (energy) that is reconstructed with good resolution. Furthermore, the shape of the background appears to be smooth over the width of the signal, so the sidebands can be used to estimate the background. However, the mass of the recoiling particle was not specified a priori, so an excess occurring in any bin in the energy range of the search could be regarded as a potential signal. The statistical implications can be quantified – this is called the *look-elsewhere effect* and is an important consideration in searches. Because the Crystal Ball did not have a powerful tracking system with a magnet, the information on the actual decay products of the $\zeta(8.3)$ was very limited. In a subsequent data sample, these signals disappeared entirely, and other experiments failed to confirm the original observation.

A second example, also from 1984, is a measurement from the UA1 experiment at the CERN Sp$\bar{p}$S collider. The paper, *Associated production of an isolated large-transverse-momentum lepton (electron or muon), and two jets at the CERN $p\bar{p}$ collider* [34], presented evidence for events consistent with the decay sequence

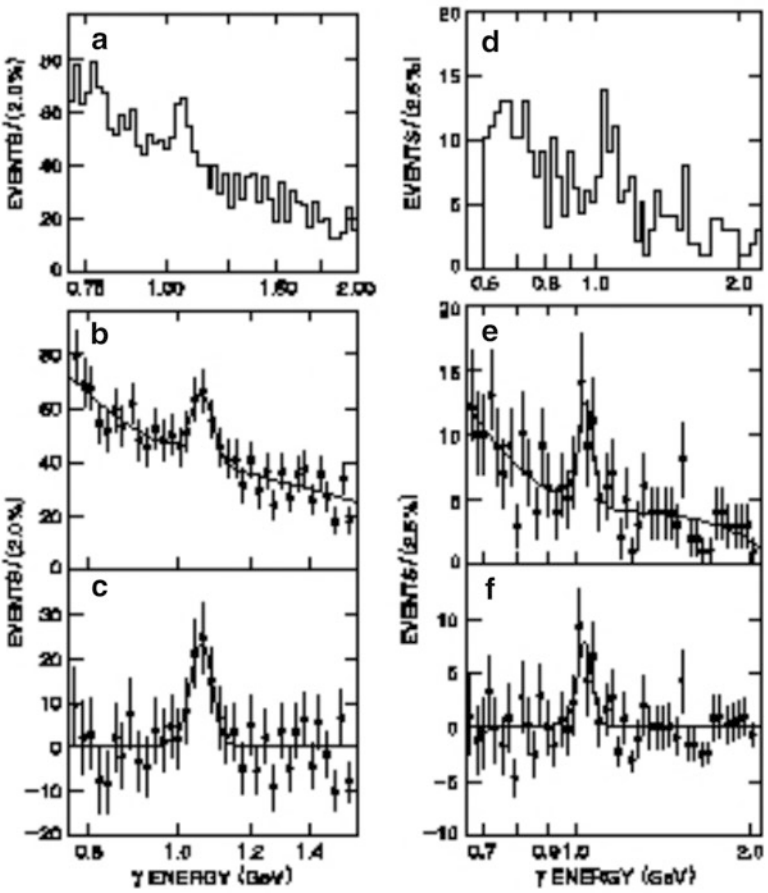

Fig. 2 Measurements from the Crystal Ball experiment in 1984, giving a preliminary indication of a resonance $\zeta(8.3)$ produced in the process $\Upsilon(1S) \rightarrow \gamma\zeta(8.3)$. The $\zeta(8.3)$ was considered by some to be a Higgs-boson candidate. The figures on the *left* show the photon-energy spectrum in the $\zeta \rightarrow$ multi-hadrons channel (**a**) without the fit to the data, (**b**) with the fit to the data (**b**), and (**c**) after background subtraction. The figures on the *right* in (**d**)–(**f**) correspond to the $\zeta \rightarrow 2$ jets channel (From Ref. [32])

$W^+ \rightarrow t\bar{b}, t \rightarrow b\ell^+\nu$. (Remember – this was back in the days when the top quark had not yet been discovered, and in this search it appeared that $m(t) < m(W)$! How would our searches at the LHC be affected if $W^+ \rightarrow t\bar{b}$?) The published paper includes kinematic distributions (Fig. 3) for two key mass combinations that appear to agree with this W-decay hypothesis. The invariant mass of the system consisting of the two highest energy jets, the lepton, and the neutrino is expected to peak at the mass of the W-boson. (Only the components of the neutrino momentum vector transverse to the beam axis are used, because the initial momenta of the colliding partons along the beam axis is unknown.) The invariant mass distribution of the

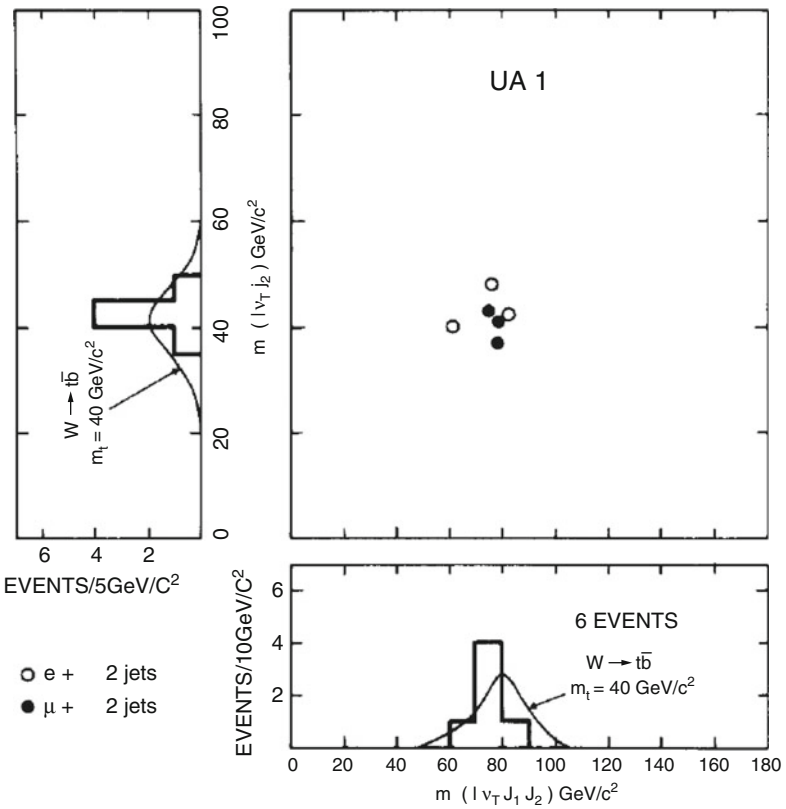

Fig. 3 UA1 experiment: kinematic distributions associated with the study of $p\bar{p}$ events with b-tagged jets, leptons, and missing transverse energy. The six events observed in data peaked both in $M(\ell\nu_T J_1 J_2)$ and in $M(\ell\nu_T J_2)$, where ν_T is the missing transverse momentum. In this study, the top quark, with mass $M_t \approx 40\,\text{GeV}$, was thought to be lighter than the W boson, so the hypothetical decay sequence was $W^+ \to t\bar{b}$, $t \to b\ell^+\nu$ (From Ref. [34])

second highest energy jet, the lepton, and the neutrino is also seen to peak at a common mass, consistent with three-body top-quark decay $t \to b\ell^+\nu$. With only six events, however, these distributions have limited power. The paper acknowledges that "that more statistics are needed to confirm these conclusions and the true nature of the effect observed." Further studies in UA1 with additional data showed that the top-quark hypothesis was not correct, and eventually the top quark was discovered at the Tevatron at a mass of around 170 GeV.

We have considered two examples of analysis problems, one from an e^+e^- experiment and one from a hadron-collider experiment. The number of signal events was quite small in both cases, which is of course common in discovery situations. As a consequence, it can be difficult to perform meaningful cross checks on the behavior of distributions. This is particularly difficult if there is a substantial

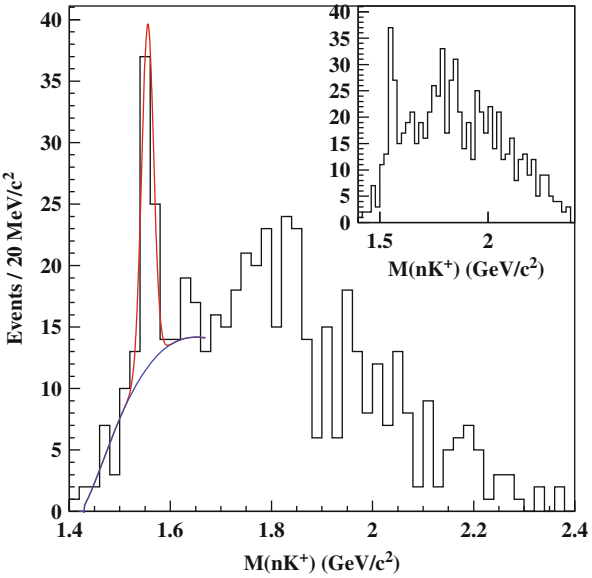

Fig. 4 CLAS experiment: study of the invariant mass distribution of the nK^+ system produced in photoproduction $\gamma d \to K^+K^-pn$. The statistical significance of the peak around 1.54 GeV was quoted to be $(5.2 \pm 0.6)\sigma$ (From Ref. [35])

background under the signal peak. Another feature shared by these searches is that the location of the signal bin in the kinematic distribution was not a priori known.

The next example, the apparent observation of pentaquark states, is truly astonishing in its scope. These hypothetical particles would have valence quark content of four quarks and one anti-quark (or the conjugate). One of the several new particles apparently found was the Θ^+ pentaquark, whose quark content was assigned to be $uudu\bar{s}$, so that a natural decay mode was $\Theta^+ \to n(udd)K^+(u\bar{s})$. Note that this particle would have baryon number $+1$, but the $\bar{s}$ gives it the opposite strangeness of a normal baryon (such as $\Lambda \sim uds$), making it "exotic."

Figure 4 shows the reconstructed invariant-mass spectrum for the nK^+ system from the photoproduction process $\gamma d \to K^+K^-pn$ in the CLAS experiment [35]. A narrow peak was observed around 1.5 GeV with a statistical significance quoted as $(5.2 \pm 0.6)\sigma$. Remarkably, nine experiments obtained evidence for a particle around this mass, each with a significance of over 4σ and several with a significance over 5σ [36, 37]. A perhaps telling sign of trouble was that the masses of these different observations were not entirely consistent.

This wave of discoveries, which even included a charm pentaquark, $\Theta_c^0 \to D^{(*)}p$, was followed by a wave of non-confirmations, and later on by a few additional positive sightings. The excitement that had begun around 2002 was dying down by 2005. Over 550 theoretical papers were produced during this period! An illuminating review of the various results, "The Rise and Fall of Pentaquarks

in Experiments," was presented by R. Schumacher [38] at the *Particles and Fields International Conference* in 2005. His review includes a comprehensive chronology of observations and non-observations of pentaquark states. He concludes that

> Thus, one can conclude that a "bandwagon" rush of over-optimistic positive sightings was in effect initially, but now the lack of convincing evidence for narrow exotic pentaquarks is overwhelming.

For many of the pentaquark searches the underlying physics of the background processes (and hence the background shapes) was not well understood. In this context, the true statistical significance of peaks can be very difficult to assess [37].

4.2 Lessons Learned: Common Problems in Searches for New Physics

What lessons can we learn from these (and many other) examples in our field? First of all, searches are difficult! Here is a list of some common mistakes or situations that occur. Do any of these affect your analysis?

- The detector may not be correctly calibrated or aligned, leading to mismeasured objects in events.
- Limitations in the detector design or technology can produce spectacular mismeasurements such as E_T^{miss} or lepton isolation in rare circumstances. Event displays can be useful for identifying unusual problems, but they can also be used in a problematic way to reject events without a well-defined procedure.
- Trigger efficiencies (including their kinematic dependence) may not be fully accounted for and can bias yields in the signal or control regions.
- Changes in the experimental conditions or calibrations may not be fully taken into account. For example, at the LHC, the presence of multiple pp collisions within a single beam crossing leads to multiple vertices and can affect many reconstructed quantities. This effect is luminosity dependent.
- A prescription for a "standard" analysis method or reconstructed object (b-tagged jets, leptons, etc.) may not give the correct result when applied in the sample of events used in your analysis. Was the standard recipe validated in an event sample in which the relevant properties are similar to yours?
- Monte Carlo event samples may not have been generated correctly.
- Monte Carlo event samples may not have correctly modeled the true physics. For example, the number of extra jets from initial- or final-state radiation may not be correct. The simulation may not model all of the kinematic correlations in the signal, leading to an incorrectly estimated signal efficiency.
- The yield in signal region can be biased by tuning selection requirements on the signal region in the data.
- The yield in the signal region can be biased by tuning selection requirements on the region used to determine the background to be subtracted.

- The background shape or normalization may be estimated incorrectly. Background estimates are especially tricky if there are contributions from many sources or if control samples are obtained with different triggers.
- Understanding the background in one kinematic region does not necessarily mean that you understand it in another region. The background composition may vary substantially from a control sample to a signal region, and the kinematic distributions may also vary between these regions.
- The shapes used in a fit may not be adequate to describe the data, which can easily produce a bias in the extracted signal yield. This effect is especially worrisome in multidimensional fits, where the shapes may not fully track the correlations among kinematic variables.
- Theoretical assumptions used to determine the backgrounds or their uncertainties may be incorrect. Consultation with theorists can be valuable in such cases.
- Systematic uncertainties may be underestimated or incomplete.
- Correlations may not be taken into account correctly. Correlations can arise from many different mechanisms. Two kinematic quantities can become correlated not only analytically, within a given sample of events, but also through a variation in the sample composition as one variable is changed.
- Backgrounds peaking under the signal may not be fully taken into account.
- The signal efficiency may be incorrectly determined.
- The signal significance may not be estimated correctly.
- The look-elsewhere effect may not have been taken into account in assessing the statistical significance.
- A signal can be created artificially as a "reflection" of a background process that produces a peak or other structure in a related kinematic variable.
- Averaging multiple measurements can be tricky; all uncertainties and their correlations must be understood.
- Bug in your program. Bug in someone else's program. Bug in ROOT.
- Advisor is in a hurry! Need to finish thesis! No time to look for more problems!
- People sometimes stop looking for mistakes or declare a result ready to be presented publically when they obtain a "desirable" result. In precision measurements, people sometimes prefer to obtain agreement with previous results, leading to a clustering of measurements that is better than the uncertainties should typically allow.
- A superposition of several of the above effects.

How many of these have you actually seen in practice? Based on conversations with students at this school, I conclude that graduate students are quite familiar with these problems, as well as many others not listed. A fundamental problem, which may simply be a statement about entropy, is that there are many ways to do something wrong, but far fewer ways to do things right! One approach, blind analysis [39, 40], offers some valuable methods but also some potential problems, especially when the event sample has not previously been explored. In general, it is important to design your analysis with as many crosschecks and control sample studies as you can to provide comprehensive tests of the analysis methods.

5 Characteristics of Standard Model Backgrounds

Before discussing specific searches for new physics, we consider in Sect. 5.1 the main SM processes that typically contribute to the backgrounds. Section 5.2 focuses on the properties of $t\bar{t}$ events, which are the dominant source of background in many new physics searches.

5.1 Survey of SM Backgrounds and Their Role in New Physics Searches

Detailed studies of SM background processes are valuable and often essential for searches. As the LHC luminosity increases, and we search for new physics with lower cross sections, the number of relevant SM background processes is increasing. These processes are interesting in their own right. If you are studying a SM process, you are contributing to the searches for new physics as well.

People sometimes believe that it is "conservative" to overestimate the background, because one is then less likely to claim a false signal. Overestimating the background is not a good practice, however. First of all, if a signal is present, you want to know it, not hide it! But even if no excess is observed, and you are setting an upper limit, subtracting an overestimated background from the yield in the signal region leads to its own problems. You will then underestimate the number of events in the signal region that can potentially be attributed to signal, and your limit will be more stringent than it deserves to be. It is not conservative to overestimate the background!

Figure 5 shows some of the key cross sections for W+jets, Z+jets, $t\bar{t}$, single-top, and diboson production processes. The cross sections for W and Z production include the branching fractions for $W \to \mu\bar{\nu}$ and $Z \to \mu^+\mu^-$, respectively. The separate production cross sections for W^+ and W^- are not shown, but these are different (as are their kinematic distributions), reflecting the charge asymmetry of the pp initial state.

Although they are not shown, the cross sections for QCD multijet processes are very large and depend strongly on the jet p_T thresholds that are applied. In fact, these cross sections are so large that it is often impossible to generate a sufficient number of Monte Carlo events to study their contribution to an LHC data sample. Although QCD multijet events can often be suppressed to a level well below that of the other backgrounds, the residual contribution must still be quantified with reliable uncertainties. Because the accuracy of QCD simulations is questionable, Monte Carlo samples are best used to gain insight into the behavior of the backgrounds rather than to determine any quantitative result. In general, QCD multijet backgrounds should be determined using control samples in the data.

Figure 6 focuses on the SM process with smaller cross sections and adds a few SUSY models for comparison. Models LM0 and LM1 are low-mass SUSY models [41] that were used in early CMS searches. They are defined in

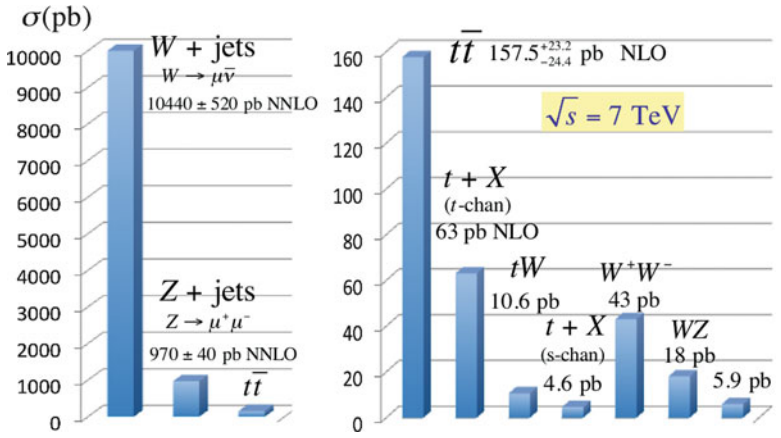

Fig. 5 Cross sections at $\sqrt{s} = 7$ TeV for common SM processes relevant to searches. Note the change in scale between the *left-* and *right-hand parts* of the figure. The cross sections for W and Z production include the branching fractions for the leptonic decay modes specified

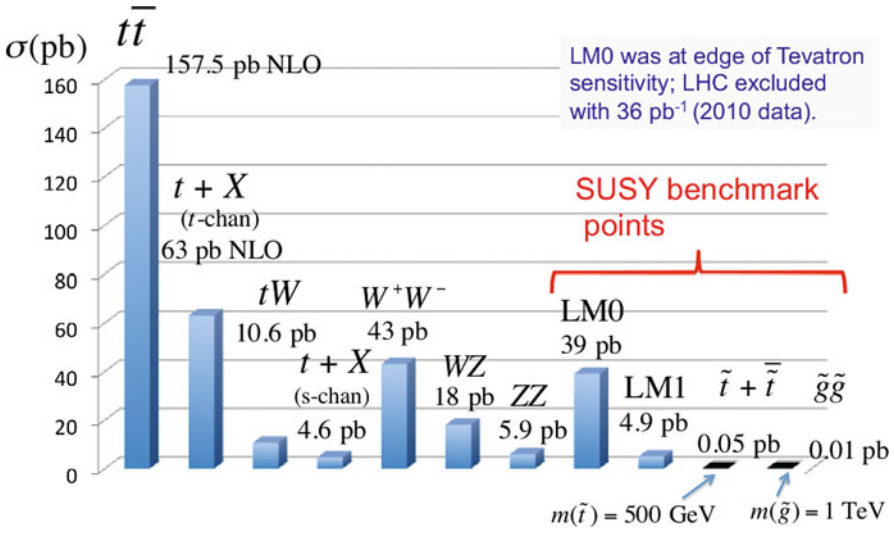

Fig. 6 Cross sections at $\sqrt{s} = 7$ TeV for processes involving top quark production and diboson production, as well as for some benchmark SUSY models

the framework of the constrained minimal supersymmetric standard model [11] (cMSSM, closely related to minimal supergravity, or mSUGRA). LM0 was defined as a reference near the edge of Tevatron sensitivity and was quickly excluded in the first LHC run. The cross sections for gluino pair production and stop (scalar top) production are also shown for certain mass parameters. These cross sections are strong functions of the SUSY particle masses and will be discussed later.

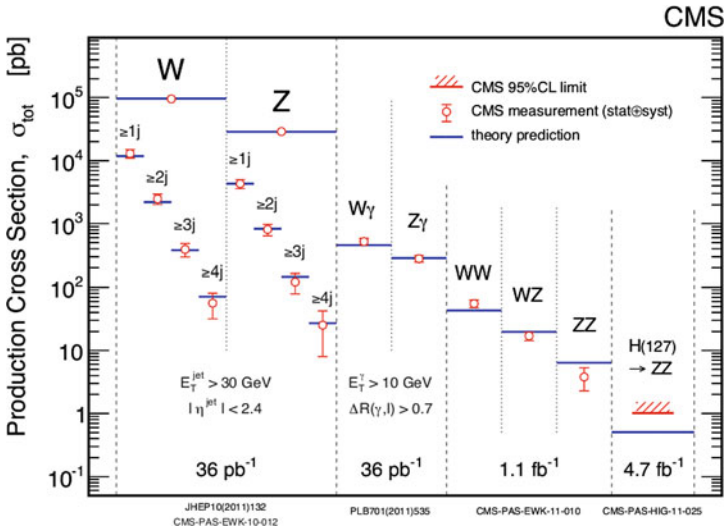

Fig. 7 Measured and theoretical cross sections for processes involving W and Z bosons ($\sqrt{s} = 7$ TeV) [42]. The cross sections fall off exponentially as the number of jets increases. The jet transverse-energy thresholds are specified in the figure

The cross sections for processes involving the production of W and Z bosons are shown in more detail in Fig. 7 [42]. The W and Z cross sections are given as a function of the number of recoiling jets.

What general observations can we make regarding the behavior of backgrounds in different search channels? Here are a few:

- Backgrounds from W+jets and Z+jets events fall off rapidly as the number of jets increases. For signatures with large numbers of jets, this effect often suppresses these backgrounds below that from $t\bar{t}$ production.
- In searches that require large missing transverse energy (E_T^{miss}), backgrounds from $W \to \ell\bar{\nu}$, $Z \to \nu\bar{\nu}$, and $t\bar{t}$ events with $W \to \ell\bar{\nu}$ typically play a major role. The large E_T^{miss} in such events is genuine, associated with high-momentum neutrinos from W or Z decay. However, mismeasured jets (sometimes associated with detector problems) or jets containing a neutrino from semileptonic b-quark decay can also lead to large E_T^{miss}. *Thus, a large value of* E_T^{miss} *does not necessarily indicate the presence of a weakly interacting particle produced in the initial hard scattering or in the decay of heavy particles.*
- In searches for signatures with large E_T^{miss} and no leptons, $(Z \to \nu\bar{\nu})$+jets represents an "irreducible" background, with the same topology as the signal. However, SM processes in which the lepton from W decay is missed can also be substantial. Such missed leptons arise not only from detector inefficiencies, but also from leptons that fail to satisfy lepton p_T and isolation requirements. Another important source of events with E_T^{miss} arises from τ-lepton decays, either to lighter charged leptons or final states with hadrons, such as $\tau^- \to \pi^-\nu$.

- In searches that include jets, E_T^{miss}, and a single isolated lepton in the signature, backgrounds arise mainly from $t\bar{t}$ and $W \to \ell\bar{\nu}$. By requiring $M_T(\ell\bar{\nu}) > 100\,\text{GeV}$ (where the ν is inferred from E_T^{miss}), backgrounds with a single W boson can be strongly suppressed, so that contributions from $t\bar{t}$ dilepton events become dominant.
- In searches that include two opposite-sign isolated leptons in the signature, $t\bar{t}$ is a critical background. If the signature involves same-flavour leptons only, the unlike-flavour sample in the data can, with care, be used to measure the contribution from $t\bar{t}$ events.
- Searches for signatures with like-sign leptons are special, in the sense that these are highly suppressed in SM processes. In $t\bar{t}$ events, a primary lepton (from W decay) and a secondary lepton (from b decay) can produce like-sign lepton background, but this contribution can be strongly suppressed with a lepton isolation requirement.
- QCD backgrounds are strongly suppressed by requiring either an isolated lepton or large E_T^{miss}. In addition, the E_T^{miss} in such events is usually aligned with one of the jets. However, the QCD cross sections are so large that one must determine whether unusual event configurations are contributing to the signal region. Such events can arise from detector mismeasurements, producing fake E_T^{miss}, or semileptonic decays of b- and c-hadrons in jets.
- Lepton isolation is a critical variable for determining whether leptons are produced in the decay of a heavy particle. However, isolation does not provide a perfect separation of such primary leptons from secondary leptons.
- In searches with one or more high-transverse-energy photons, the isolation of the photon plays a critical role, similar to that for leptons.
- A requirement of multiple b jets, which is applied in many searches for processes with $\tilde{t}$ or $\tilde{b}$ squarks, helps to suppress $W+$jets and $Z+$jets backgrounds. This behavior is another reason why $t\bar{t}$ is such an important background process.
- New physics processes with rates comparable to those from these SM common processes are now largely excluded. As a consequence, searches for new physics typically require a careful understanding of the *tails* of the kinematic distributions of SM processes. As the luminosity increases, additional rare SM processes will become relevant.

Methods for determining background contributions range from simple to highly involved. Regardless of the method, it is always important to understand the background composition and to explore how it varies as the selection criteria are applied. Simulated event samples are extremely useful for this purpose. If the signal is a sufficiently narrow peak over a slowly varying background, the background is usually estimated from a fit the effectively extrapolates the sidebands into the signal region. In many searches, including nearly all searches for SUSY, the signal is simply an excess of events in the tail of a distribution such as E_T^{miss}. In this case, much more effort and care is required to obtain a reliable background prediction.

In the simplest approach, simulated events samples are generated, reconstructed, and analyzed using procedures that are as close as possible to those used for the data.

Typically, corrections must also be applied to account for known differences between the actual detector and the simulated detector. Because the trigger conditions typically vary over the data-taking period (for example, as the luminosity increases), it is difficult to model them correctly in simulated event samples. To simplify the determination of the trigger efficiency, one typically applies an offline selection requirement that is somewhat more stringent than the most stringent trigger requirement, establishing a uniform condition over the full running period. It is also common practice to set the offline requirement such that the trigger efficiency is on the plateau with respect to the applied thresholds.

The use of simulation for predicting backgrounds has several potential problems, which are widely recognized. Because searches for new physics processes typically involve event-selection requirements that strongly suppress SM backgrounds, the amount of residual background often depends on the how the backgrounds behave in a narrow region of phase space. The modeling of the so-called tails of the kinematic distributions may not be as accurate as the modeling of the cores, where most of the events are, and where the simulation is often validated most fully with control samples. In addition, some types of detector problems may not be modeled in simulated event samples. It is not unusual to see the quantitative agreement between data and simulation worsen significantly as the analysis cuts are applied.

In practice, one rarely sees an analysis in which the key background estimates are obtained simply by taking the yields from simulation, normalized to the integrated luminosity of the data sample. A more common practice is to normalize a distribution from simulation either in a sideband region that should be relatively free of signal (for the model considered!) or in a control region obtained by altering one or more of the cuts. This procedure has some virtues, especially that the burden on simulation is much reduced. However, it is not entirely free of potential problems; for example, the composition of the control sample may not be fully understood. It can also be difficult to reliably quantify the uncertainty on the scale factors required to translate the observed background yield in the control region to the observed background yield in the signal region.

The term *data-driven background prediction* is used to describe any method that relies largely on control samples in the data to estimate the background. An example is the use of a photon + jets control sample to predict the background from $Z \to \nu\bar{\nu}$+jets events, a highly non-trivial exercise. The best data-driven methods rely on specific, well-understood properties of SM processes for which the uncertainties can be quantified in a well-defined manner.

5.2 Discussion of a Key SM Background: $pp \to t\bar{t}$

Many searches involve high-mass objects, which have complicated decay chains and signatures containing a large number of jets and other objects. (A notable exception is the search for new heavy gauge bosons using the signatures $Z' \to \ell^+\ell^-$

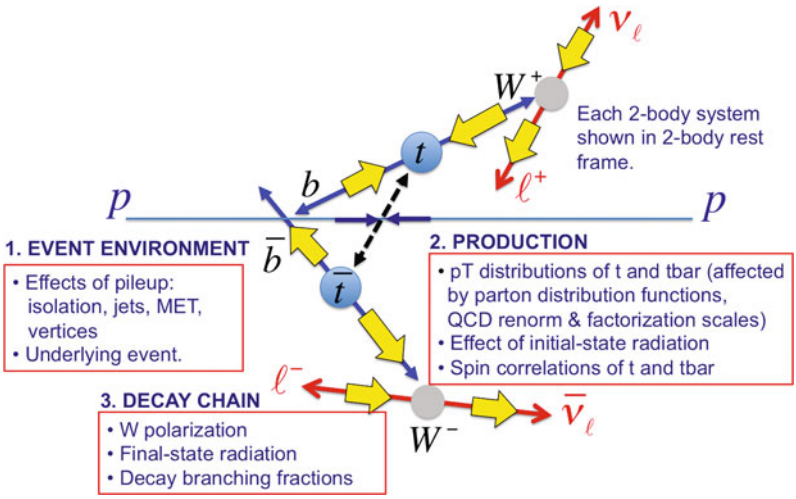

Fig. 8 An overview of key issues in understanding $t\bar{t}$ production in pp collisions: event environment, production, and decay chain

and $W'^- \to \ell^-\bar{\nu}$.) Because of the large top-quark mass, SM $t\bar{t}$ production is a prototypical background, leading to events with high jet multiplicity, isolated leptons, and E_T^{miss}.

Figure 8 summarizes the key experimental issues that arise in $t\bar{t}$ events. These can be divided into three (somewhat arbitrarily defined) categories: (1) the event environment, (2) the production properties, and (3) the decay chains. The *event environment* encompasses such features as multiple pp collisions and the properties of the underlying event (the particles that are not produced in the hard scattering processes). *Production* effects include the p_T distribution of the top quarks. These are affected strongly by the parton distribution functions, of course, but the p_T distributions are more directly relevant to measurements. The extent of the tails of the p_T distributions can be particularly important for searches involving E_T^{miss}, because $t\bar{t}$ events with the very highest E_T^{miss} values usually arise when neutrinos from $t \to bW^+$, $W^+ \to \ell^+\bar{\nu}$ are Lorentz boosted to high energy in the laboratory frame. The production of additional jets from initial- and final-state radiation can be an important issue for analyses in which jet multiplicity plays a key role. For example, a SUSY search in the dilepton final state might well require the presence of at least three jets to suppress background from $t\bar{t}$ in which both W bosons decay leptonically. The decay chains produce only two (b) jets, one each from $t \to bW$, but QCD radiation can produce additional jets. Finally, the *decay chain* itself involves effects such as the W-boson spin polarization, which controls the angular distribution of the W-boson decay products and hence their momenta in the laboratory frame.

As an example of one of these issues, let's consider the W-boson spin polarization in top-quark decay and its effect on the decay $W^+ \to \ell^+\nu$, where ℓ^+ is

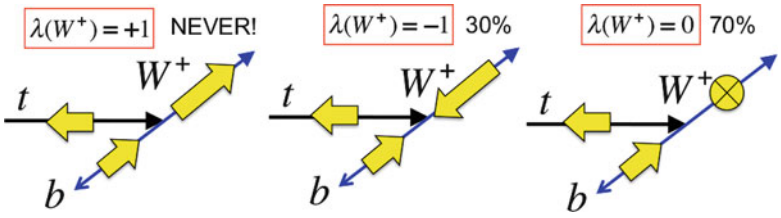

Fig. 9 Examples of spin configurations in top-quark decay. The top quark is shown (arbitrarily) in the helicity state $\lambda(t) = -1/2$; this is not important for the discussion. Because of the $V - A$ coupling at the decay vertex, the daughter b-quark, which is relativistic, is predominantly in the state $\lambda(b) = -1/2$. The W^+ boson cannot then be in the state $\lambda(W) = +1$ because this would yield an angular momentum projection along the decay axis of $3/2$, which is greater than the spin of the t quark

any charged lepton. This sequential two-body decay process is well understood in the SM, and it has been studied experimentally, although interesting new physics effects could in principle enter at a low level. As discussed below, the effects of QCD on the W-boson polarization have been calculated to NNLO; these corrections are small with respect to the basic, weak-interaction behavior. Here are some questions to think about:

1. In the top-quark rest frame, which distribution is harder, the momentum spectrum of the charged lepton or the neutrino?
2. Is the polarization of the W boson the same in t and $\bar{t}$ decay? Hint: no, but they are directly related.
3. Are the kinematic distributions of the lepton and neutrino the same for t and $\bar{t}$ decay? Hint: yes (fortunately).

Figure 9 shows three spin configurations for the decay of a top quark. We begin with the top quark, which is shown in the laboratory frame in a helicy $\lambda = -1/2$ state, indicated by a fat yellow arrow pointing in the direction opposite to its momentum vector. This does not mean that top quarks can only be produced with this helicity; in fact, in strong production there is no preference for either helicity. (There are, however, correlations between the helicities of the two top quarks, which can generate small but noticeable effects in a dilepton analysis.) The figure also shows the decay $t \to bW^+$, illustrated with back-to-back momentum vectors for the b-quark and the W^+ boson. *These momentum vectors are shown in the t-quark rest frame*. Thus, the drawing shows two *different* reference frames, the lab frame and the t-quark rest frame. This convention is commonly used because it breaks the analysis of the decays into two parts: (1) angular distributions in the rest frame of the decaying particle and (2) Lorentz boosts to the frame in which the decaying particle is observed.

Top-quark decay is controlled by a $V - A$ coupling at the tW^+b vertex, which couples only to the left-handed chiral projection of the b quark. *In the relativistic limit,* this left-handed chiral projection maps onto the helicity $\lambda(b) = -1/2$

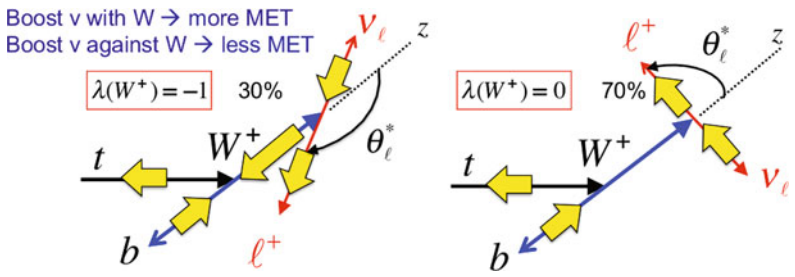

Fig. 10 The sequential two-body decay process $t \to bW^+$, $W^+ \to \mu^+\bar{\nu}$. *Each two-body decay is shown in its respective rest frame.* For the case $\lambda(W^+) = -1$ the lepton is emitted preferentially in the backward direction in the W^+ rest frame. The neutrino is correspondingly emitted preferentially in the forward direction, creating an asymmetry between the lepton p_T and E_T^{miss} distributions in the t-quark rest frame and in the laboratory frame

component of the b-quark. Because $m_b << m_t$, the b-quark is in fact relativistic, so the amplitude for helicity $\lambda(b) = -1/2$ completely dominates, and in each of these three cases shown in Fig. 9, the b-quark is shown with this helicity.

The decay configuration on the left, however, is forbidden by conservation of angular momentum. Because the momenta of the b-quark and the W^+ boson are aligned in the t-quark rest frame, we can sum all of the angular momenta along this axis. For the configuration with $\lambda(W^+) = +1$ and $\lambda(b) = -1/2$, the magnitude of the total angular momentum along the decay axis must be $3/2$. This spin projection would be greater than the spin of the top quark (1/2), so it cannot possibly conserve angular momentum. Note that *there cannot be any orbital angular momentum projection along a two-body decay axis*, because $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. Any *orbital* angular momentum must be *perpendicular* to this axis! Thus, to a very good approximation, there are only two allowed helicities for the W^+ boson: $\lambda(W^+) = 0$, which it turns out occurs about 70 % of the time, and $\lambda(W^+) = -1$, which occurs 30 % of the time. There is a tiny amplitude for $\lambda(W^+) = +1$ and $\lambda(b) = +1/2$, which is present because the b quark is not massless. These probabilities are reliable SM predictions and are calculated to be $f_0 = 0.687 \pm 0.005$, $f_{-1} = 0.311 \pm 0.005$, and $f_{+1} = 0.0017 \pm 0.0001$ [43] at NNLO in QCD.

We turn now to the decay of the W^+ boson into $\ell^+\nu$, where the W boson is produced in top-quark decay (Fig. 10). For the case $\lambda(W^+) = -1$, the lepton is emitted preferentially in the backward direction, with a distribution given by $dN/d\cos\theta_{\ell^*} \sim (1-\cos\theta_\ell^*)^2$. This result uses conservation of angular momentum, which implies that in any two-body decay $A(J, M) \to B(\lambda_B) + C(\lambda_C)$, the distribution of the polar angle θ of particle B with respect to the z axis is given by [44, 45]

$$dN/d\cos\theta \sim [d^J_{M,\lambda(B)-\lambda(C)}(\theta)]^2. \tag{3}$$

Here, J is the spin of the parent particle A and M is its spin projection along the z axis; $\lambda(B)$ is the helicity of particle B; and $\lambda(C)$ is the helicity of particle C. Thus, for $\lambda(W^+) = -1$, the neutrino is emitted in the forward direction with respect to the momentum of the W^+ as observed in the t-quark rest frame; it is therefore boosted to higher energy in that frame. For $\lambda(W^+) = 0$, the lepton angular distribution is symmetric in θ_ℓ^*: $dN/d\cos\theta_\ell^* \sim \sin^2\theta_\ell^*$. For this W polarization, the distributions of the lepton and neutrino momentum in the top-quark rest frame are the same.

The helicities of the W, lepton, and neutrino all reverse when we switch from t to $\bar{t}$ decay, leading to the result that the angular distributions for the lepton are the same in t and $\bar{t}$ decay chains. (The probability for $\lambda(W^-) = +1$ is $\approx$30 % in $\bar{t} \to \bar{b}W^-$ decay, corresponding to the probability for $\lambda(W^+) = -1$ in $t \to bW^+$ decay.) Thus, for both t and $\bar{t}$ decay, the neutrino (E_T^{miss}) distribution is harder than that of the lepton p_T distribution in the laboratory frame. The relationship between the lepton spectrum and the E_T^{miss} spectrum in $t\bar{t}$ events has been used as the basis for data-driven background predictions in a number of SUSY searches [46,47].

Other examples of phenomenology papers that provide important information on SM processes relevant to new physics searches are the predictions for W-boson polarization fractions in $W+$jets events [48] and predictions for $\gamma+$ jets events and their relationship to $Z+$ jets events [49].

6 Searches for Supersymmetry at the LHC

Supersymmetry (SUSY) is a general framework that encompasses many different theories or models, which are associated with the specific mechanisms that break the symmetry. Each of these models can have a broad range of parameter values. Thus, the most generic approach to SUSY leads to many distinct phenomenological situations, presenting challenges for both experiment and theory. An issue of special importance is defining the set of criteria used to trigger the readout of the detector. There are interesting models for which, without special care, the detector would not even trigger on SUSY events. The large number of models also creates a challenge in *interpreting* the results of a given search, because exclusion plots for one model often cannot be translated into limits for a different model.

Theorists have developed models in which the number of parameters is reduced by applying various constraints. For example, the constrained Minimal Supersymmetric Standard Model (cMSSM) [11, 50, 51] has just four continuous real parameters and a sign, but many theorists do not consider its underlying assumptions to be especially well motivated. A more generic approach is the phenomenological minimal supersymmetric standard model (pMSSM) [52, 53], which incorporates a number of phenomenological constraints and has 19 real parameters beyond those of the SM. Another recent theoretical strategy, exemplified by *simplified models* [54–57], has been to focus on distinct phenomenological signatures that can be interpreted in more than one theory. In Sect. 6.1 we begin with a basic introduction to SUSY phenomenology. Section 6.2 describes the methods used in some of the important searches and summarizes their results.

6.1 A First Look at SUSY Phenomenology

SUSY is based on a mapping between fermionic and bosonic degrees of freedom. A SM spin-1/2 particle, such as the electron, has two spin states, so that two matching bosonic degrees of freedom are required. When a SUSY transformation is performed on a SM field, the transformed field has the same gauge quantum numbers as the original field: each of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ quantum numbers is exactly preserved. For example, when a SUSY transformation acts on a gluon field, yielding a gluino field, the gluino has exactly the same colour quantum numbers as the gluon and hence transforms under $SU(3)_C$ rotations in the same way, according to the adjoint representation.

Returning to leptons, the $SU(2)_L$ quantum numbers of the electron are different for the left- and right-handed chiral projections. The e_L is part of an $SU(2)_L$ weak-isospin doublet together with the electron neutrino, ν_e. The e_R, in contrast, transforms as a singlet under $SU(2)_L$ rotations: it has zero weak isospin and does not couple to the W boson. Because SUSY preserves these quantum numbers, each of these chiral projections is a degree of freedom of the electron that maps onto its own scalar electron, or selectron. These scalar partners are designated as $\tilde{e}_L$ and $\tilde{e}_R$, even though they themselves are spinless. The subscripts mean that the selectrons are the *partners* of the left- and right-handed electrons; furthermore, they have the corresponding L and R electroweak gauge quantum numbers. Similarly, the L- and R-handed chiral components of each quark map separately onto two scalar quarks (squarks), $\tilde{q}_L$ and $\tilde{q}_R$. In general, the SUSY partners $\tilde{f}_L$ and $\tilde{f}_R$ of an SM fermion f have different masses after SUSY breaking.

In the minimal supersymmetric standard model (MSSM), the Higgs sector requires two complex doublet fields, not just the one we are familiar with in the SM. (The MSSM is discussed extensively at this school by Sven Heinemeyer.) Altogether, the MSSM has 124 free parameters, including the 18 parameters of the SM embedded within it. We have already discussed how the fermion fields in the SM (associated with leptons and quarks) map onto scalar fields (associated with sleptons and squarks), and how the gluon field maps onto the gluino field. This leaves the electroweak gauge bosons and the Higgs bosons.

Figure 11 lists the particles that make up the electroweak gauge and Higgs sectors of the MSSM. The left-hand table lists the gauge and Higgs bosons, while the middle table lists their fermionic SUSY partners, the gauginos and higgsinos. In each case, there are a total of 16 degrees of freedom. In general, mixing effects among the neutral gauginos and neutral higgsinos lead to the set of physical particles (mass eigenstates), the neutralinos, which are designated by the symbol $\tilde{\chi}_i^0$. Similarly, mixing effects among the charged gauginos and charged higginos lead to physical particles called charginos, $\tilde{\chi}_i^{\pm}$. These particles are listed in the table at the right of Fig. 11. There are four neutralinos ($\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, $\tilde{\chi}_4^0$), numbered in order of increasing mass. There are four charginos, $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^{\pm}$. Each of these "-ino" particles has spin-1/2. As before, there are 16 degrees of freedom. Both the gluinos and the neutralinos are Majorana fermions. If you think it is crazy to more or less double the number of particles, consider the prediction of antimatter!

Particle	J	Degrees of freedom
W^+	1	3
W^-	1	3
Z	1	3
γ	1	2
H	0	1
h	0	1
H^+	0	1
H^-	0	1
A	0	1
Total		16

Particle	J	Degrees of freedom
$\tilde{W}^+$	1/2	2
$\tilde{W}^-$	1/2	2
$\tilde{Z}$ | $\tilde{W}^0$	1/2	2
$\tilde{\gamma}$ | $\tilde{B}$	1/2	2
$\tilde{H}$	1/2	2
$\tilde{h}$	1/2	2
$\tilde{H}^+$	1/2	2
$\tilde{H}^-$	1/2	2
Total		16

Mixing

Particle	J	Degrees of freedom
$\tilde{\chi}_1^+$	1/2	2
$\tilde{\chi}_1^-$	1/2	2
$\tilde{\chi}_2^+$	1/2	2
$\tilde{\chi}_2^-$	1/2	2
$\tilde{\chi}_1^0$	1/2	2
$\tilde{\chi}_2^0$	1/2	2
$\tilde{\chi}_3^0$	1/2	2
$\tilde{\chi}_4^0$	1/2	2
Total		16

Fig. 11 The gauge and Higgs sectors of the Minimal Supersymmetric Standard Model (MSSM). The table on the *left lists* the gauge and Higgs bosons of the MSSM, which together have 16 degrees of freedom. The MSSM requires two complex Higgs doublets, not just one, as in the case of the SM. The SUSY partners, gauginos and higgsinos, are listed in the *middle table*. Mixing among the neutral gauginos and higgsinos leads to mass eigenstates called neutralinos; mixing among the charged gauginos and higgsinos leads to mass eignenstates called charginos, shown in the *right-hand table*

SUSY, if it exists, must be a broken symmetry because partners with masses equal to those of the SM particles would already have been discovered. (This fact does not compromise the SUSY solution to the gauge-hierarchy problem as long as the SUSY breaking mechanism is *soft*, as discussed in Ref. [11].) SUSY breaking is a complex subject with various scenarios; this phenomenon occurs in a so-called *hidden sector* of particles that have no tree-level interactions with the *visible sector* (e.g., the MSSM spectrum discussed earlier). The breaking of SUSY is then transmitted from the hidden to the visible sector through some mediation mechanism, which can be a set of additional particles constituting a *messenger sector*. The proposed mechanisms include gravity-mediated SUSY breaking [11, 58], leading to a heavy gravitino ($\tilde{G}$), and gauge mediation, leading to a very light gravitino, with mass typically in the eV range. In models with gauge mediation [59–67] the next-to-lightest SUSY particle (NLSP) can decay into its superpartner plus a gravitino, for example, $\tilde{\chi}^0 \to \gamma\tilde{G}$, $\tilde{\chi}^0 \to Z\tilde{G}$, or $\tilde{\tau}_R^{\pm} \to \tau^{\pm}\tilde{G}$. Whatever the SUSY breaking mechanism, SUSY particles still have the same SM gauge properties as their ordinary SM partners. This is a key point when thinking about the phenomenology of the decay modes. Your intuition from the SM will serve you surprisingly well!

The MSSM possesses $B - L$ symmetry, which leads to a multiplicatively conserved quantum number called R-parity [68],

$$R = (-1)^{3(B-L)+2S}, \tag{4}$$

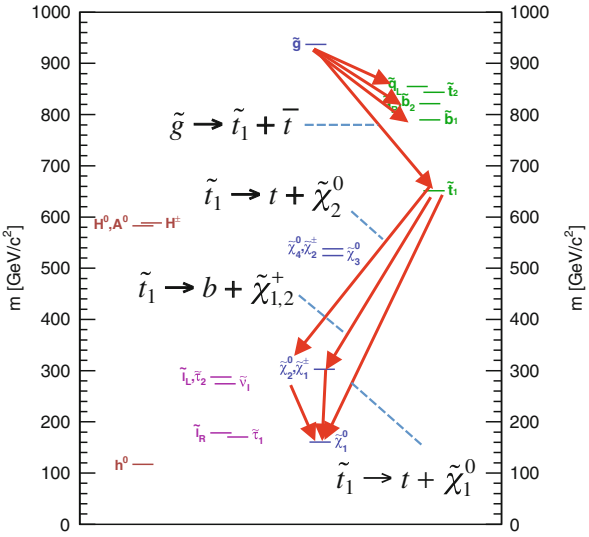

Fig. 12 Mass spectrum for the benchmark model LM6, with some of the possible decay modes indicated. These correspond to the processes $\tilde{g} \to \tilde{t}_1 \bar{t}$, $\tilde{t}_1 \to t\tilde{\chi}_{1,2}^0$, and $\tilde{t}_1 \to b\tilde{\chi}_{1,2}^+$. In pp collisions, any of the SUSY particles can be produced directly, although particles with colour charge typically have larger cross sections

where B is the baryon number, L the lepton number, and S the spin of the particle. You can verify that for all ordinary SM particles $R = 1$, while for all the SUSY partners $R = -1$. A valid fundamental vertex of the SM can be converted into a valid fundamental vertex involving SUSY particles by replacing an even number of SM particles with their SUSY partners. Conservation of R-parity has major consequences:

1. Starting from an initial state containing only SM particles, SUSY particles must be produced in pairs.
2. The decay chain of a SUSY particle must end with the production of the lightest supersymmetric particle (LSP), which is stable, and which in many scenarios is $\tilde{\chi}_1^0$. Because it is stable and only weakly interacting, the $\tilde{\chi}_1^0$ LSP is a potential dark-matter candidate. Events in such models are typically characterized by large E_T^{miss} resulting from the presence of two such SUSY decay chains, each ending with a $\tilde{\chi}_1^0$.

Searches for models without R-parity conservation cannot rely on the E_T^{miss} signature and are typically quite different in their strategy.

Figure 12 shows the mass spectrum of the model LM6, which has been used as a benchmark by CMS but is now in the excluded part of the cMSSM parameter space. In this model, the gluino is the heaviest SUSY particle, while $\tilde{\chi}_1^0$, the lightest neutralino, is the LSP. In the case of the stop ($\tilde{t}$), large mixing can arise between the L- and R-handed SUSY partners ($\tilde{t}_L$ and $\tilde{t}_R$), resulting in a large mass splitting

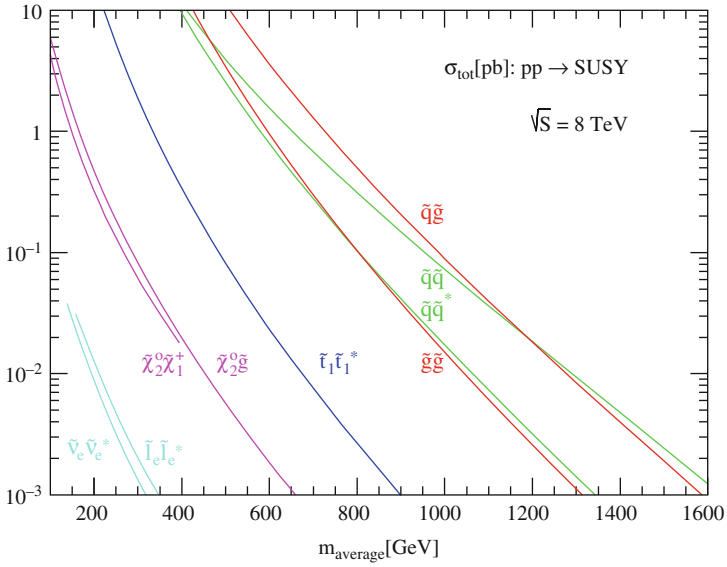

Fig. 13 Cross sections for SUSY particle production at $\sqrt{s} = 8$ TeV, based on Prospino. Particles with colour (squarks and gluinos) can be produced strongly and (for a given mass) have much larger cross sections than particles that can only be produced through electroweak interactions. However, the decay signatures of particles produced via electroweak processes can be quite distinctive, allowing sensitive searches to be performed. In this figure, an *asterisk* denotes an antiparticle, not an off-shell particle (Figure courtesy of Tilman Plehn)

between the mass eigenstates. (See, for example, Ref. [13].) These particles are labeled $\tilde{t}_1$ (lighter) and $\tilde{t}_2$ (heavier). From Fig. 12 it is clear that in the LM6 model, $\tilde{t}_1$ is substantially lighter than all of the other squarks, followed by $\tilde{b}_1$.

The phenomenology of a given SUSY model can often be understood in a reasonably straightforward way from the mass spectrum and mixing parameters, together with the usual gauge couplings. The two key issues for experimental searches are the production cross sections and the decay branching fractions. We consider these in general and then return to the example of LM6.

The production cross sections for SUSY particles at $\sqrt{s} = 8$ TeV as a function of their masses are shown in Fig. 13. The particles fall into two broad categories: those with colour charge (squarks and gluinos), which can be produced via the strong interactions, and those that have only have electroweak couplings: the sleptons, sneutrinos, charginos, and neutralinos. The large cross sections for strongly produced particles represent a big advantage for searches, but some of the particles produced via electroweak processes can produce very distinctive signatures, making the searches quite feasible, even at very low cross sections. In this plot, the symbol $\tilde{q}$ represents the sum over $\tilde{u}$, $\tilde{d}$, $\tilde{c}$, $\tilde{s}$, and $\tilde{b}$, with both L- and R-handed partners included. The cross section for production of $\tilde{t}_1\bar{\tilde{t}}_1$ is much less than that for $\tilde{g}\tilde{g}$ at the same mass (and is also much smaller than that

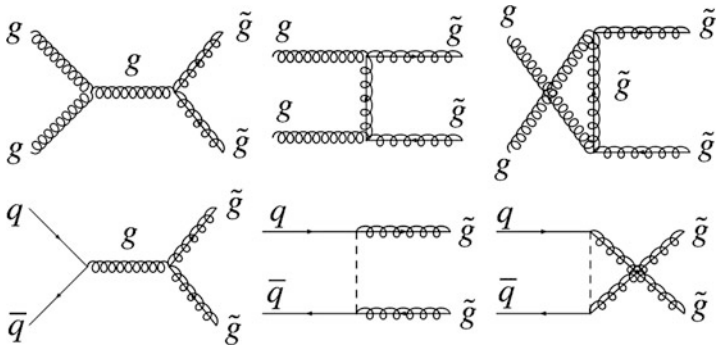

Fig. 14 Feynman diagrams leading to the production of a pair of gluinos in pp collisions in an R-parity-consverving SUSY model. The *dashed lines* in the intermediate state denote squarks

for $t\bar{t}$, as discussed in Sect. 6.2). SUSY particle production at the LHC, including uncertainties from parton distribution functions and other sources, is discussed in Refs. [69–71].

Figure 14 shows the diagrams for processes contributing to gluino pair production in a SUSY model with R-parity conservation. The decay of a gluino proceeds in analogy to the SM process $g \to q\bar{q}$, governed by the same strong coupling constant. In an R-parity conserving model there are four possible decay modes for each quark flavour:

$$\tilde{g} \to q + \bar{\tilde{q}}_L, \qquad \bar{q} + \tilde{q}_L, \qquad q + \bar{\tilde{q}}_R, \qquad \bar{q} + \tilde{q}_R. \tag{5}$$

Two cases arise:

1. $m(\tilde{g}) > m(\tilde{q}) + m(q)$: true two-body decay
2. $m(\tilde{g}) < m(\tilde{q}) + m(q)$: the squark is virtual (three-body decay)

The subscripts L, R (or $1, 2$) have been omitted for generality. In the case of three-body decay, an example of a decay chain with a virtual $\tilde{b}$ squark is

$$\tilde{g} \to \tilde{b}_i^* \bar{b}, \qquad \tilde{b}_i^* \to b\tilde{\chi}_1^0, \tag{6}$$

where the $\tilde{b}^*$ indicates a virtual squark and i denotes either L, R or $1, 2$. (Other possible squark decays are discussed below.) This decay sequence involves both a strong and a weak interaction vertex and leads to

$$\tilde{g} \to b\bar{b}\tilde{\chi}_1^0 \to \text{jet} + \text{jet} + E_T^{\text{miss}}. \tag{7}$$

Of course, if the two-body decay is allowed, the branching fraction for the three-body mode is highly suppressed.

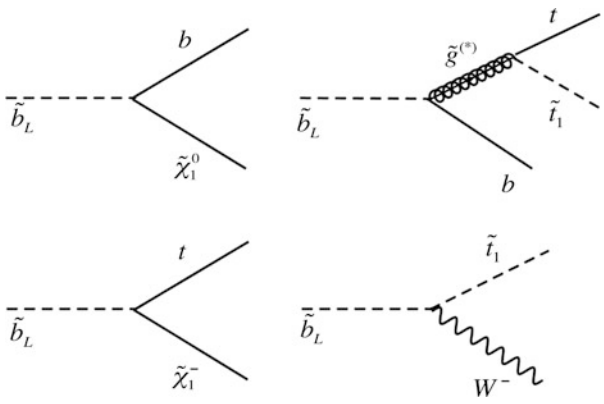

Fig. 15 Examples of diagrams for the decays of squarks (scalar quarks). The relative importance of the different processes depends on the particle masses, as well as on the mixing parameters that determine the gaugino/higgino content of the charginos and neutralinos

Here is a simple question: How many gluino decay modes are there in LM6? For each of five flavours (u, d, c, s, and b) there are four modes (see Eq. 5). However, for top, there are only the two modes $\tilde{g} \to \tilde{t}_1 \bar{t}$ and $\tilde{g} \to \bar{\tilde{t}}_1 t$, because $\tilde{t}_2$ is too heavy to be produced together with a t-quark. Thus, there are a total of 22 gluino decay modes in this model.

In SUSY models motivated by naturalness considerations, the $\tilde{t}_{L,R}$ and $\tilde{b}_L$ are typically constrained to be light, while the gluino is *not too heavy*. In pp collisions, stops can be produced in two main ways: (i) directly, via pair-production processes such as $gg \to \tilde{t}\bar{\tilde{t}}$, and (ii) indirectly, via gluino pair production $gg \to \tilde{g}\tilde{g}$, with $\tilde{g} \to \tilde{t}\bar{t} + \bar{\tilde{t}}t$. Because the production cross section for $\tilde{t}\bar{\tilde{t}}$ (or $\tilde{b}\bar{\tilde{b}}$) is much smaller than that for $\tilde{g}\tilde{g}$ (at the same mass), gluino pair production is a potentially useful way to search for stop and sbottom. Note also that, even if the $\tilde{b}_1$ mass turns out to be larger than that of $\tilde{t}_1$, the combined mass of the particles in the $\tilde{t}\bar{t}\tilde{t}\bar{t}$ final state could still be comparable to that in the $\tilde{b}\bar{b}\tilde{b}\bar{b}$ final state. There are many different possibilities to consider! We will discuss these scenarios in more detail in Sect. 6.2. Finally, we note that the $\tilde{u}$ and $\tilde{d}$ squarks, for which the corresponding quark flavour u and d is found in the proton, can be produced in additional processes. Inclusive hadronic searches (without b-tagging) place constraints on these squarks.

Several paradigms for squark decay are shown in Fig. 15. A squark can decay both via neutralino emission, as in $\tilde{b}_L \to b\tilde{\chi}_1^0$, and via chargino emission, as in $\tilde{b}_L \to t\tilde{\chi}_1^-$. Referring back to Fig. 12, let's consider the possible decays of $\tilde{t}_1$ and $\tilde{t}_2$ in LM6. The lighter mass eigenstate, $\tilde{t}_1$ has four possible decay modes: $\tilde{t}_1 \to t\tilde{\chi}_1^0$ (25 %), $\tilde{t}_1 \to t\tilde{\chi}_2^0$ (16 %), $\tilde{t}_1 \to b\tilde{\chi}_1^+$ (43 %), and $\tilde{t}_1 \to b\tilde{\chi}_2^+$ (16 %). The $\tilde{t}_2$ is significantly heavier and has four additional decay modes: $\tilde{t}_2 \to t\tilde{\chi}_3^0$, $\tilde{t}_2 \to t\tilde{\chi}_4^0$, $\tilde{t}_2 \to \tilde{t}_1 h$, and $\tilde{t}_2 \to \tilde{t}_1 Z$.

6.2 Examples of Searches for Supersymmetry

A typical SUSY search begins with a set of topological requirements for an appropriate set of reconstructed objects such as jets (or b-tagged jets), leptons, photons (especially in gauge-mediated SUSY), and E_T^{miss}. The term E_T^{miss} is confusing, because energy is not a vector and therefore cannot have a transverse component. This quantity is the magnitude of the missing transverse momentum vector, $E_T^{\text{miss}} = |\mathbf{p}_{\text{T}}^{\text{miss}}|$, where the *missing transverse momentum* is given by

$$\mathbf{p}_{\text{T}}^{\text{miss}} = -\left[\sum_{i=\text{objects}} \mathbf{p}_{\text{T}}^{i}\right]. \tag{8}$$

This calculation of $\mathbf{p}_{\text{T}}^{\text{miss}} = \mathbf{p}_{\text{T}}^{\text{init}} - \mathbf{p}_{\text{T}}^{\text{observed}}$ uses conservation of momentum and the fact that the initial-state momentum transverse to the beam direction is known to be zero to a very good approximation. (The energy label originates from the use of calorimeter measurements, which are important because the contributions of neutral particles, both photons and neutral hadrons, must be included.) An analogous calculation cannot be performed in the direction along the beams (z direction) because the colliding partons each carry unknown fractions (x_1 and x_2) of the proton momenta. If sufficiently well measured, E_T^{miss} and $\mathbf{p}_{\text{T}}^{\text{miss}}$ can be attributed to unobserved final-state particles.

The details of how the sum in Eq. 8 is performed over reconstructed objects are important. The objects can be jets above some minimum p_T threshold; in this case the variable is usually called MHT rather than E_T^{miss}. In an E_T^{miss} calculation, the objects are often calorimeter cells (both electromagnetic and hadronic) or calorimeter towers, combining different parts of the calorimeter that point back to the interaction point. In CMS, a particle flow technique is used in which information from the tracker and calorimeter is carefully combined to improve the resolution.

Figure 16 shows the E_T^{miss} resolution in ATLAS [72] as a function of another key global event variable, $\sum E_T$ (event), which is the *scalar sum* over the transverse momenta of the jets above threshold. ATLAS has used a variety of different control samples to measure the resolution, yielding consistent results that roughly follow a $\sqrt{\sum E_T}$ dependence (see figure). For backgrounds with large E_T^{miss} from neutrinos (such as leptonic $t\bar{t}$ events), the precise shape of the E_T^{miss} resolution function is usually not critical, because most of the E_T^{miss} in the event is genuine. But if there is substantial background from sources with fake E_T^{miss} (such as QCD multijet events), the effects of non-gaussian tails of the E_T^{miss} resolution function must be more carefully quantified. In CMS, a variety of specific instrumental effects that generate fake E_T^{miss} have been identified, and software filters have been developed to suppress such events.

The scalar sum of the jet transverse momenta (above some threshold, typically in the range 30–50 GeV) is usually denoted H_T (rather than $\sum E_T$),

$$H_T = \sum_{i=\text{jets}} p_T^i, \tag{9}$$

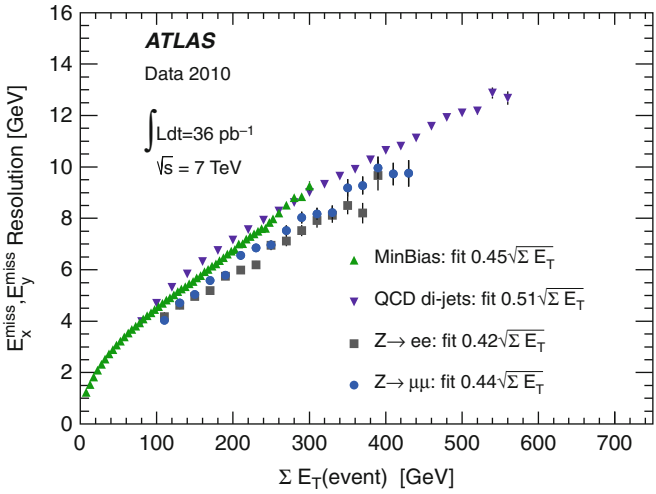

Fig. 16 Measured resolution on E_T^{miss} in the ATLAS experiment, using several different control samples. The results are consistent with $\sqrt{\sum E_T}$ dependence (From Ref. [72])

and is another discriminating variable commonly used in SUSY searches. Other interesting variables used in SUSY searches are α_T (discussed below), M_{T2} [73,74], and the razor variables [75].

From the discussion in Sect. 6.1, it is clear that many SUSY production and decay scenarios can arise, and that a finely tuned optimization for each one is unwieldy. Partly for this reason, the initial SUSY searches performed by ATLAS and CMS were *inclusive*, based on simple topological signatures. These searches can be regarded as surveys to determine whether the event yields in the main channels are consistent with SM expectations. The main inclusive search topologies are

- Jets + E_T^{miss} (all-hadronic search; veto events with observed leptons)
- 1 lepton (e or μ) + jets + E_T^{miss}
- 2 leptons + jets + E_T^{miss} (same-sign or opposite-sign leptons)
- 1 photon + jets + E_T^{miss}
- 2 photons + jets + E_T^{miss}
- $\geq$3 leptons + jets + E_T^{miss}

The like-sign dilepton channel and the trilepton channel are special in that SM backgrounds are highly suppressed. Although the number of expected signal events is typically very small for relevant SUSY models, the sensitivity can still be quite high. In the opposite-sign dilepton channel one can include a Z-boson selection. In all cases above, b-tagging can be used to define a search in a subsample of events with increased sensitivity to $\tilde{t}$ or $\tilde{b}$ decays.

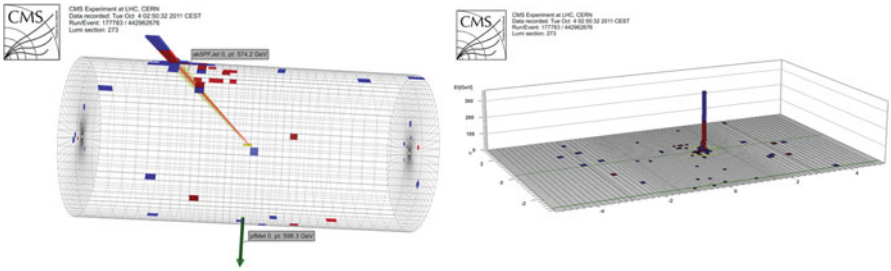

Fig. 17 Event display for a monojet event in the CMS experiment, showing the energy desposited in the electromagnetic calorimeter (*red*) and the hadronic calorimeter (*blue*). There are many monojet events in the data sample, but this does not mean that SUSY has been discovered!

6.2.1 SUSY and Dark-Matter Searches in Monojet Final States

The first search that we will consider, however, has the amusing signature of $pp \rightarrow$ nothing, more or less. This is the simplest possible search–just look for nothing! This final state could correspond, for example, to the production of a pair of neutralinos, $pp \rightarrow \tilde{\chi}_i^0 \tilde{\chi}_j^0$. This is a weak process, so the cross section is small. For the case $i = j = 1$, this process is related to $\tilde{\chi}_1^0 p \rightarrow \tilde{\chi}_1^0 p$, which effectively corresponds to a (cosmic) direct dark-matter search. But while there is a strong motivation, it is obvious that one cannot perform this search quite as described. There is, however, a beautiful method, based on the fact that one can trigger on events with initial-state radiation, either a gluon or a photon. Both ATLAS [76, 77] and CMS [78, 79] have performed such searches. The collision after the radiation occurs can then proceed as before, but at a somewhat lower center-of-mass energy. Such processes lead to *monojet events*, which in fact are straightforward to find in the data. Figure 17 shows a monojet event from CMS. A highly energetic jet is recoiling against nothing, so there is a large amount of E_T^{miss}.

Does this mean that we have discovered SUSY? As always, the question is, what are the backgrounds? Unfortunately, the process $pp \rightarrow Z + 1$ jet, $Z \rightarrow \nu\bar{\nu}$ produces monojet events, and there is a smaller background from $W + 1$ jet as well. (It is possible to suppress $t\bar{t}$ and QCD multijets backgrounds to a very low level.) The Z background is measured in data by scaling the yields from a $Z \rightarrow \mu^+\mu^-$ control sample. Figure 18 shows the distribution of E_T^{miss} from an ATLAS monojet search. Searches by both ATLAS and CMS have produced remarkably sensitive results. With some care, such LHC results can be translated into the type of dark matter exclusion plots (cross section vs. WIMP mass) obtained from direct dark-matter detection experiments [80, 81].

Even if an excess with respect to the SM were observed, one should not jump to the conclusion that SUSY is the explanation. Many signatures for new physics can admit more than one explanation. The "problem" of identifying which kind of new physics is the actual source of an observed excess is the kind of problem we want to have!

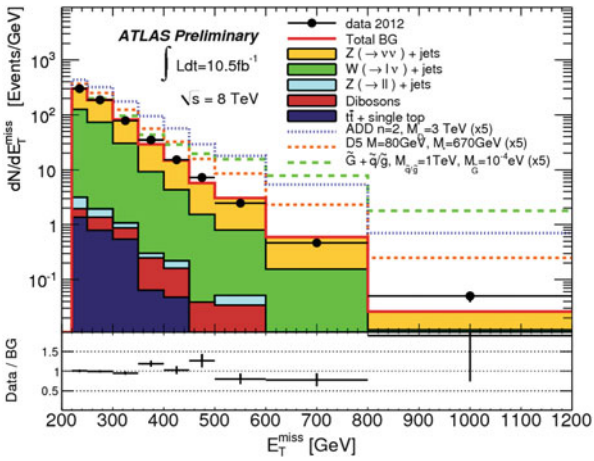

Fig. 18 ATLAS monojet search: observed distribution of E_T^{miss} in data (points with error bars), together with expected SM backgrounds (stacked histograms) and expectations from various signal hypotheses. Note that the dominant background arises from Z + jets, with $Z \to \bar{\nu}\nu$. The x-axis of the plot begins at 200 GeV (From Ref. [76])

6.2.2 SUSY Searches in All-Hadronic Final States

Searches in the dijet + E_T^{miss} or multijets + E_T^{miss} channel are sensitive to production of SUSY particles via strong interactions. For example, the production of a pair of squarks can lead to a dijet + E_T^{miss} final state via the process $pp \to \tilde{q}\bar{\tilde{q}}$ with $\tilde{q} \to q\tilde{\chi}_0^1$. This final state has an enormous background from QCD dijet events. The jets in such background events, however, are typically back to back with equal energies. A variable that measures these characteristics is [82],

$$\alpha_T = \frac{p_T^{j_2}}{M_T(j_1, j_2)} = \frac{\sqrt{p_T^{j_2}/p_T^{j_1}}}{\sqrt{2(1-\cos\Delta\phi)}}, \tag{10}$$

which has been used to dramatically suppress the QCD dijet background [83]. (It has also been generalized to treat multijet events by forming two pseudo-jets.) Here, j_1 and j_2 are the first- and second-leading jets in p_T and $\Delta\phi$ is the angle between them in the transverse plane. Well-measured QCD dijets events are balanced ($\alpha_T = 0.5$), while SUSY events such as $\tilde{q}\bar{\tilde{q}}$ production often have $\alpha_T > 0.5$. *Mismeasured* QCD events usually have $\alpha_T < 0.5$.

Several strategies have been used to study the multijets + E_T^{miss} channel. Although Z + jets and $t\bar{t}$ are the dominant backgrounds at high E_T^{miss}, it is critical to have a reliable measurement of the QCD multijet background as well. Figure 19 shows the H_T and missing H_T distributions from a CMS search [84] in this channel, after the application of a basic set of preselection requirements.

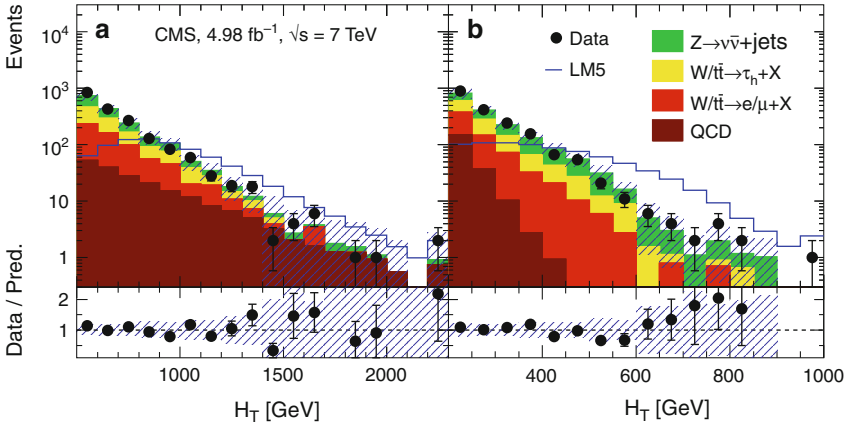

Fig. 19 CMS experiment: distributions of H_T, the scalar sum of jet p_T values, and missing H_T in a search for SUSY in the multijets + E_T^{miss} channel. The SM backgrounds are displayed as stacked histograms, while the expectation for the SUSY LM5 benchmark model is shown as an overlaid histogram. Because the logarithmic scale can sometimes hide discrepancies, it is useful to compare the data with the predicted SM background on a linear scale, as shown below the main histograms (From Ref. [84])

(Missing H_T is calculated using jets rather than calorimeter cells and is better suited to the data-driven QCD background estimation method used in this analysis, which involves a jet-energy-smearing procedure.) The preselection requires at least three jets with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.5$, and events are vetoed if they contain an isolated lepton with $p_T > 10\,\text{GeV}$. Because fake E_T^{miss} in QCD multijet events is usually associated with a single, badly mismeasured jet, the E_T^{miss} in such events is usually aligned with a jet; this background can therefore be suppressed with the requirement $\Delta\phi(\mathbf{p}_T^{\text{miss}}, j_1) > 0.5$, where j_1 is the leading jet. (Similar cuts are applied to the second and third leading jets.)

Both of the kinematic variables H_T and missing H_T provide sensitivity to a SUSY contribution, which is shown overlaid on the stacked histograms for the background predictions. In the missing H_T distribution, the QCD multijet contribution falls off more rapidly than the other backgrounds, which produce genuine E_T^{miss} associated with neutrinos. The contribution from $Z \to \nu\bar{\nu}$+jets is critical (see Fig. 19) and is measured with a γ+jets control sample. While $Z \to \mu^+\mu^-$+jets provides an alternative method, the small branching fraction for leptonic Z decay is a severe limitation in the statistical power of the control sample. Theoretical support for this method has been important to relate the γ+jets control sample to the Z+jets background. In this search, essentially all backgrounds are determined with data-driven methods, and the observed yields in the signal regions are consistent with the background predictions.

6.2.3 SUSY Searches in Final States with Leptons

SUSY processes generate many different signatures with leptons. In the SM, leptons can be produced in processes mediated by γ, Z, $W^{\pm}$, and Higgs bosons, but not by gluons. Because SUSY preserves gauge quantum numbers, an analogous statement holds in SUSY models. Sleptons (scalar leptons) and sneutrinos can be produced directly in electroweak processes (with small cross sections, as shown in Fig. 13), or in the cascade decays of other SUSY particles, once either electoweak gauge bosons or their SUSY partners are produced (either on- or off-shell). Figure 12 shows that squark decays can lead to the production of neutralinos or charginos. The neutralinos can decay via processes such as $\tilde{\chi}_2^0 \to \tilde{\ell}_{L,R}^{\pm}\ell^{\mp}$, $\tilde{\chi}_2^0 \to \tilde{\nu}\bar{\nu}$, $\tilde{\chi}_2^0 \to \tilde{\bar{\nu}}\nu$, $\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 h$, and $\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 Z$. The charginos can decay via processes such as $\tilde{\chi}_1^+ \to \tilde{\nu}_{\ell L}\ell^+$, $\tilde{\chi}_1^+ \to \tilde{\ell}_L^+\nu$, and $\tilde{\chi}_1^+ \to \tilde{\chi}_1^0 W^+$. When neutralinos and charginos (or W and Z bosons) are produced, leptonic signatures become important.

Decays of $\tilde{\chi}_i^0$ and $\tilde{\chi}_j^{\pm}$ give rise to some of the most famous SUSY signatures. The decay of the heavy neutralino can proceed through the cascade,

$$\tilde{\chi}_2^0 \to \tilde{\ell}^{\pm}\ell^{\mp},\ \tilde{\ell}^{\pm} \to \ell^{\pm}\tilde{\chi}_1^0, \tag{11}$$

where the first decay is analogous to $Z \to \ell^+\ell^-$. The scalar lepton $\tilde{\ell}$ can be either on- or off-shell; its decay preserves the flavour of the original lepton-slepton pair, so the two final-state leptons have opposite sign and same flavour. The distribution of invariant masses of the dilepton system is a powerful tool that is unusual in SUSY searches. Although there is no peak in this mass spectrum, it has important kinematic features, including a well-defined upper edge.

One of the most basic leptonic SUSY searches involves a signature with a single lepton, jets, and E_T^{miss}. As it happens, one of the students at this school, Jeanette Lorenz, has been closely involved in such a search on ATLAS. We consider her paper [85] as an example. The analysis requires at least four jets with $p_T > 80\,\text{GeV}$ and one isolated lepton with $p_T > 25\,\text{GeV}$. Events are vetoed if there is a second isolated lepton with $p_T > 10\,\text{GeV}$. This requirement helps to suppress $t\bar{t}$ dilepton events. Of course, SM events with a single lepton, jets, and E_T^{miss} arise from $t\bar{t}$ events or from W+jets events, with a leptonic $W \to \ell\bar{\nu}$ decay in either case. Such true single-lepton events in which both the lepton and neutrino are produced in the decay of a single W boson can be suppressed using the transverse mass quantity,

$$m_T \equiv \sqrt{2p_T^{\ell}E_T^{\text{miss}}(1-\cos\Delta\phi)}, \tag{12}$$

where the lepton mass has been ignored and ϕ is the angle between the lepton and the E_T^{miss} vector in the transverse plane. When the lepton and the E_T^{miss} in an event both arise from $W^- \to \ell^-\bar{\nu}$ decay, m_T approximates the mass of the lepton-neutrino system, and the distribution of m_T cuts off around the W boson mass. In contrast, the m_T distribution in many SUSY models extends well above the W mass because the source of E_T^{miss} is χ_1^0 production, which is effectively decoupled from the lepton.

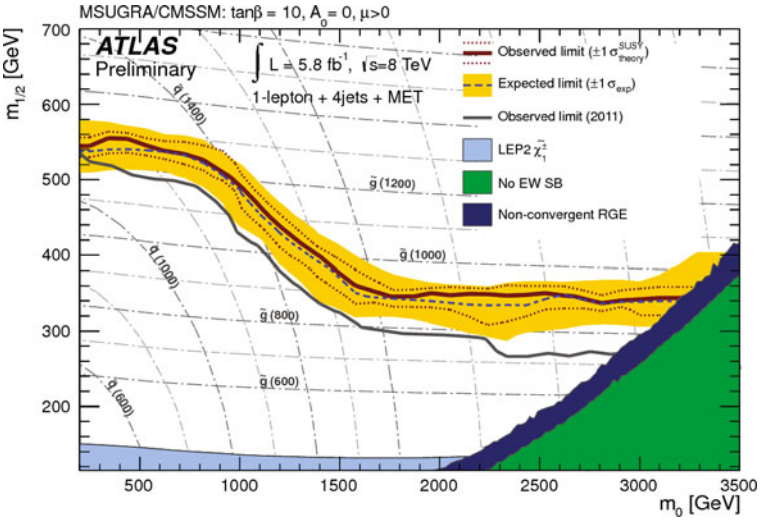

Fig. 20 Exclusion region in the $m_{1/2}$ vs. m_0 plane of the cMSSM from an ATLAS search for SUSY in the single lepton + jets + E_T^{miss} channel (From Ref. [85])

A key ingredient of Jeanette's analysis is the use of several control regions, which are used to monitor the main backgrounds. These regions are cleverly defined using b-tagging and anti-b-tagging to separate the $t\bar{t}$ and W+jets contributions. The regions are defined in a region of intermediate E_T^{miss} to suppress potential contamination from a SUSY signal. As a perspective on the degree of background rejection involved in such an analysis, I estimate from Jeanette's paper that, comparing the number of $W \to \ell\bar{\nu}$+jets events produced to the number contributing to the signal region, the rejection factor is around 3×10^{-7}, while for single-lepton $t\bar{t}$, the rejection factor is around 2×10^{-4}. These impressive factors give an idea of how SUSY searches must strongly suppress SM backgrounds. CMS results for the single-lepton final state are presented in Ref. [47].

6.2.4 Interpreting SUSY Results

Many ATLAS and CMS searches, especially the initial studies, were interpreted using the constrained Minimal Supersymmetric Standard Model (cMSSM). This framework provides a means to compare results with searches from the Tevatron and LEP. The cMSSM contains just five parameters, which are defined at the grand unification (GUT) scale: a common scalar fermion mass (m_0), a common gaugino mass ($m_{1/2}$), a common trilinear coupling (A_0), the ratio of vacuum expectation values for u- and d-type fermions ($\tan\beta$), and the sign of the higgsino mass parameter (sign μ). Figure 20 shows the excluded region in the $m_{1/2}$ vs. m_0 plane for fixed values of the other cMSSM parameters, which are specified at the top of the figure.

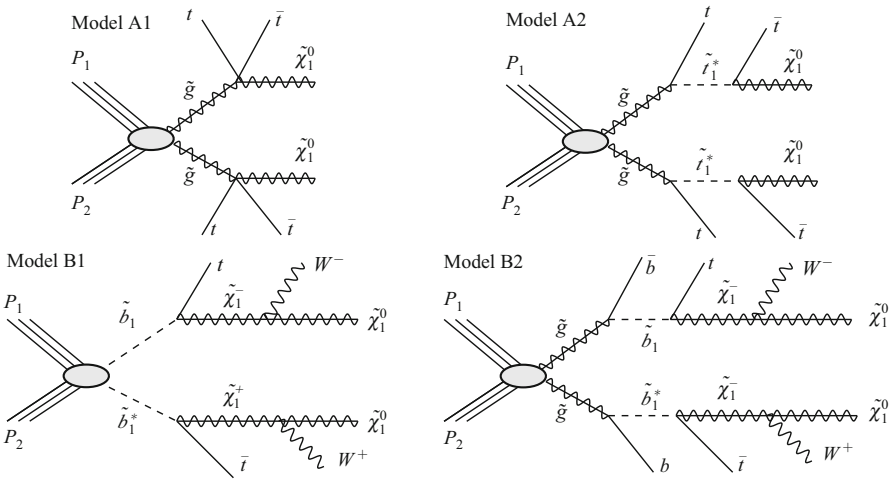

Fig. 21 Simplified models used for the CMS like-sign dileptons + b jets search. *Model A1* corresponds to gluino pair production with decays to off-shell stop; *model A2* is similar with gluino decay to on-shell stop. *Model B1* corresponds to direct production of sbottom, and *model B2* corresponds to gluino pair production followed by production of sbottom (From Ref. [88], a CMS search for like-sign dileptons with b jets. In this figure an *asterisk* (*) denotes an antiparticle, not an off-shell particle)

These plots can be mysterious, but it is useful to note that the contour lines of fixed gluino mass are nearly horizontal, with $m(\tilde{g}) \approx 2.5m_{1/2}$, while the contours of fixed squark mass have more curvature, but are roughly vertical for large values of m_0. To produce an exclusion plot of this type, it is necessary to generate simulated event samples for the signal at a grid of model points covering the parameter space of the plot. For each point, one determines whether the signal yield predicted for the given SUSY model parameters (using the predicted cross section, usually at NLO) can be excluded on the basis of the observed event yield in data, taking into account the predicted SM background. We can see that Jeanette's analysis excludes gluino masses below 0.9–1.3 TeV (roughly), depending on the value of m_0.

While the GUT-scale constraints increase the predictive power of the cMSSM (and allow us to make beautiful plots), many theorists regard these constraints with some suspicion. In addition, the constraints lead to relationships between SUSY particle masses at the electroweak scale that are not sufficiently generic to cover the ranges of all important scenarios. A gluino mass excluded in the context of a cMSSM interpretation might not be excluded in a more generic model that permits a broader range of mass splittings. Small mass splittings generally lead to less E_T^{miss} and/or softer jets, and therefore to lower signal efficiencies and poorer sensitivity.

As noted earlier, the framework of *simplified models* has been developed to provide a more generic description of relevant new physics processes. Figure 21 show several examples of (the many) models that can be defined within this

framework. Each model describes a single production and a single decay chain; a null result from a search can be used to place an upper limit on the cross section that can be associated with the full process. For a given set of mass parameters, one can also calculate a reference cross section for the simplified model that can be tested against observations in data. This allows one to exclude the given set of parameters. Note, however, that if the same mass values are embedded in a complete SUSY spectrum, the branching fractions can very well change because additional decay channels can become available. This effect can weaken the mass constraints. Simplified models have been especially useful for studies motivated by naturalness, where the number of relevant SUSY particles is typically small.

6.2.5 SUSY Searches Motivated by Naturalness

The concepts of *fine tuning* and *naturalness* were described in the introduction, and they have been discussed extensively in the literature. The discovery of a Higgs-like particle at $m \sim 125\,\text{GeV}$ has strengthened what were previously hypothetical arguments, and it is now more urgent to confront the question of whether and how the mass of this spin-0 particle is protected against enormous quantum corrections. A neutral way to formulate this question is to say that we would like to determine experimentally whether nature is fine tuned and, if not, to identify the mechanism that avoids the need for fine tuning. Models for new physics beyond the SM that avoid fine tuning are called *natural models,* and they include SUSY models with certain characteristics that we discuss here.

The paper *Natural SUSY Endures* [13] provides a useful starting point to learn about these issues. (A student at this school, C.-T. Yu, is also a co-author of a recent paper on related natural SUSY phenomenology [86].) The implications of naturalness can be found by analyzing the effects that contribute to the quadratic terms in the Higgs potential, including higher order corrections from gauge and Yukawa interactions. Naturalness can be interpreted to mean that such terms are similar in size, with magnitudes set by the electroweak scale ($v \sim 246\,\text{GeV}$). In the context of SUSY, these considerations lead to the following conclusions:

1. The masses of $\tilde{t}$ (both stops) and $\tilde{b}_L$ (but not $\tilde{b}_R$) are less than 500–700 GeV,
2. The gluino is not too heavy, below 900 GeV–1.5 TeV, and
3. The higgsinos ($\tilde{H}$) are also light, leading to one chargino and two neutralinos with masses less than 200–350 GeV. Neutralino and chargino states are designated collectively as electroweakinos or EWKinos.

The masses of the other SUSY partners do not play an important role in suppressing the Higgs quantum corrections and so are much less constrained. Their masses could be greater than 10 TeV and not affect fine-tuning considerations. As a consequence, many recent SUSY searches have focused on the states listed above. Both direct production of squark-antisquark pairs and indirect production via gluino decays are important channels; gluino pair production has a larger cross section if the gluino mass is not too large.

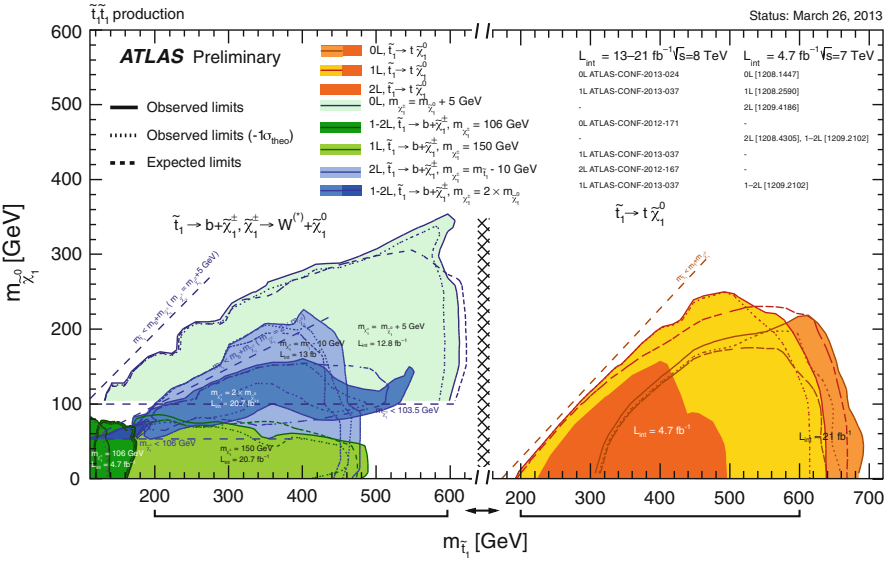

Fig. 22 ATLAS limits on direct stop production [87]. The *left-hand part* of the figure shows limits obtained from searches for $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_1^{\pm} \to W^{(*)}\tilde{\chi}_1^0$, while the *right-hand part* shows limits obtained from searches for $\tilde{t}_1 \to t\,\tilde{\chi}_1^0$

The production and decays of SUSY particles was discussed in Sect. 6.1. A key point is that the cross section for direct pair production of squarks is very small, unless their SM partners are valence quarks in the proton (see Fig. 13). Searches for $\tilde{t}\bar{\tilde{t}}$ and $\tilde{b}\bar{\tilde{b}}$ must therefore contend with small cross sections. Note that, because squarks are scalars, their direct production cross sections are suppressed relative to those for fermion pair production because there is only one spin state to sum over. The contribution to squark pair production from $q\bar{q} \to \tilde{t}\bar{\tilde{t}}$ is also suppressed near threshold by the factor β^3 (where β is the velocity of the $\tilde{t}$), because the $\tilde{t}\bar{\tilde{t}}$ must be produced in an $\ell = 1$ state (p-wave). Finally, squark production is suppressed relative to gluino production because of the different colour factors for the two cases. Besides the small cross sections, an additional challenge arises in direct-production searches: kinematically, the $t\bar{t}$ background shares many of the overall features of the $\tilde{t}\bar{\tilde{t}}$ signal.

In spite of these challenges, significant progress has been made using the $\sqrt{s} = 8$ TeV data sample to search for light stop, sbottom, and EWKinos. Figure 22 shows the results from ATLAS searches for direct stop production (incorporating updates after this school) [87]. The excluded scenarios are regions in the $m(\tilde{\chi}_1^0)$ vs. $m(\tilde{t}_1)$ plane, and are based on searches in zero lepton, one lepton, and dilepton final states. The use of b-tagging plays a major role in these searches; fortunately this tool is very well developed in both ATLAS and CMS. Figure 22 is divided into two parts, according to the stop decay channel assumed. The process $\tilde{t}_1 \to b\tilde{\chi}_1^{\pm}$,

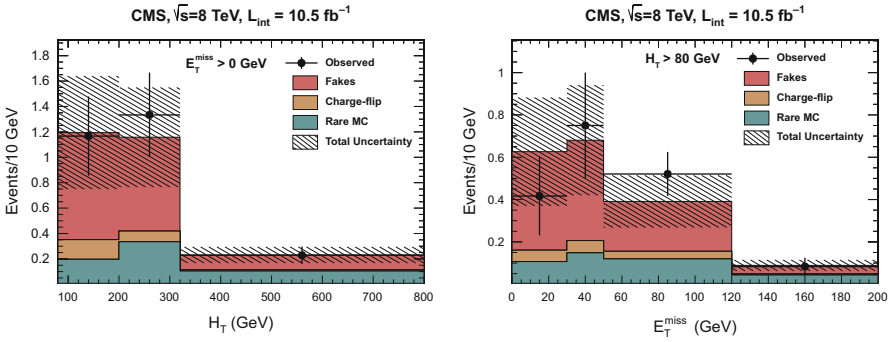

Fig. 23 CMS experiment: distributions of H_T and E_T^{miss} for events with like-sign dileptons and b-jets (From Ref. [88])

$\tilde{\chi}_1^{\pm} \to W^{(*)}\tilde{\chi}_1^0$ is assumed for the exclusion regions shown on the left, while the decay $\tilde{t}_1 \to t\chi_1^0$ is shown on the right. Sensitivity to $\tilde{t}\bar{\tilde{t}}$ production cuts off at large $\tilde{t}_1$ masses because of the corresponding fall off in cross section. Sensitivity also falls off as the $\tilde{\chi}_1^0$ mass increases, because the spectrum becomes *compressed*, resulting in small values of E_T^{miss} and softer jets. For model parameters near the diagonal in this plot, the results are sensitive to initial-state radiation, which affects the high end of the E_T^{miss} distribution. The presentation of the search results for light stop is reasonably well suited to the simplified models approach. However, the results shown in Fig. 22 incorporate several assumptions and the interested reader should refer to the original papers for more information.

Figure 23 shows results from a search for light-stop search from CMS in the final state with like-sign dileptons + b-jets [88]. Pairs of isolated like-sign leptons are rare in SM processes. They can arise from processes in which one lepton is primary, from W-boson decay, and the second is secondary, for example, from b decay. Most of the secondary leptons are not isolated (they are inside or near b-jets), but some are. In addition, effects such electron charge misidentification as a result of bremsstrahlung must also be understood. Finally, there are a small number of rare SM processes that actually produce same-sign dileptons, such as $t\bar{t}Z$ and $t\bar{t}W$. The CMS search considers events with two isolated leptons (e or μ) with $p_T > 20\,\text{GeV}$, at least two b-tagged jets with $p_T > 40\,\text{GeV}$, and large E_T^{miss}. Figure 23 shows the distributions of H_T and E_T^{miss} for the events satisfying these criteria: the data are consistent with the background predictions. (The ATLAS results are presented in Ref. [89].) As we noted before in the context of monojet searches, a signal in a final-state such as like-sign dileptons would not point to a unique source of new physics. In fact, the absence of any signal so far has been used to establish limits on several SUSY scenarios, including sbottom pair production, gluino pair production with off-shell stops, and gluino pair production with on-shell stops.

Let's briefly consider pair production of neutralinos and charginos (EWKinos). Because these particles do not have colour charge, the cross sections are generically

much smaller than those for gluinos and squarks. On the other hand, the signatures can be distinctive, and it is possible that the EWKino masses are small, boosting their cross sections. We have already discussed the famous neutralino cascade process $\tilde{\chi}_2^0 \rightarrow \tilde{\ell}^{\pm}\ell^{\mp}$, $\tilde{\ell}^{\pm} \rightarrow \ell^{\pm}\tilde{\chi}_1^0$ that gives rise to a pair of opposite-sign, same flavour leptons. An even more distinctive signature is that of trileptons, which can be produced in processes such as $pp \rightarrow \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$. SM processes rarely produce three isolated primary leptons; the analysis must therefore carefully measure the contribution from events with at least one "fake lepton" (which can in fact be a lepton from semileptonic b-quark decay).

7 Exotica Searches

This section describes examples of searches that have novel features, either in their methodologies or in their physics goals. These examples highlight the fact that the range of possibilities at the TeV scale is vast, and we must try to investigate as many of them as possible.

7.1 Search for Large Extra Dimensions

As noted in the introduction, SUSY is not the only approach to resolving the hierarchy problem: both extra spatial dimensions and technicolour provide alternatives. In fact, the monojet search discussed in connection with a search for neutralino pair production can also be interpreted in the context of models of extra dimensions. How is that? Let's start with the model of Arkani-Hamed, Dimopoulos, and Dvali (ADD) [25], which provides a completely different approach to the hierarchy problem from that of SUSY. This model postulates that in a fundamental sense, there is no difference between the weak scale and that of gravity, once gravity is properly understood. The idea is that gravity appears to us to be extremely weak (and the associated Planck scale M_{Pl} appears to be correspondingly very high, creating the huge difference with respect to the electroweak scale) because gravity (and not the other forces) propagates in additional dimensions besides those that we observe. In Large Extra Dimensions (LED) models, n extra spatial dimensions of size R are postulated; the "true" Planck scale in $4 + n$ dimensions is given by $M_D^{2+n} = M_{\mathrm{Pl}}^2/R^n$, which can be made compatible with the electroweak scale by making R sufficiently large. (The hierarchy problem is then translated into a new question about why R or n is so large.)

The LED hypothesis has motivated challenging measurements of gravity at sub-millimeter distance scales. Furthermore, because the true, higher dimensional gravity is strong at the electroweak scale, it should be possible to produce gravitons in LHC collisions. The compactification of the extra dimensions results in a Kaluza-Klein "tower" of massive graviton excitations. Because the gravitons propagate in

the extra dimensions, they can escape detection, leading to an E_T^{miss} signature similar to that from SUSY models. Studies of monojet and monophoton events have yielded limits that exclude values of M_D below ~3 TeV for n in the range 2–6 [76–79].

A more modern but related idea, the Randall-Sundrum (warped extra dimension) model [24], requires just one extra dimension. In this model, the extra dimension separates two $3 + 1$ dimensional surfaces (branes) in the full higher-dimensional space. Gravity is concentrated on one brane, while the particles of the SM reside on the other. Gravity can propagate in the bulk region between the two branes, but it is exponentially attenuated. It is this attenuation that makes gravity appear weak, rather than the dilution effect that operates in LED models. As in the case of LED, a Kaluza-Klein tower of graviton modes is produced, but in this case the decay signature does not involve large E_T^{miss}. Gravitons can be produced in $q\bar{q}$ or gg s-channel processes and then decay into pairs of SM particles, including photons [90].

7.2 *Search for Long-Lived Stopping Particles*

Imagine a particle that lives long enough that it does not decay during the beam crossing interval when it was produced, but simply stops somewhere in the detector and eventually decays. Such particles are predicted in a variety of different scenarios, including hidden valley models [91, 92] and models with split supersymmetry [93, 94]. Let's consider a split SUSY scenario, in which the gluino and neutralino (LSP) have masses at the LHC energy scale but all of the scalar SUSY particles are at some extremely high mass scale. The gluino then has a long lifetime, because the two-body decay $\tilde{g} \to \tilde{q}\bar{q}$ is forbidden. Possible decays are $\tilde{g} \to g\tilde{\chi}_1^0$, which must proceed via a loop diagram since the neutralino has no colour, and $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$. A process to compare in the SM is the β decay of a free neutron, $n \to pe^-\bar{\nu}_e$, which leads to the neutron lifetime of about 10 min.

What happens to a long-lived gluino? As a consequence of its long lifetime, it hadronizes into an object called an R-hadron, which can be $\tilde{g}g$, $\tilde{g}q\bar{q}$, $\tilde{g}qqq$, and so on. The R-hadron interacts with the material in the detector and, some fraction of the time, will stop, typically in the densest region. Figure 24 (left) shows a map of these regions in CMS, while Fig. 24 (right) shows a simulated R-hadron decay.

The first question to ask ourselves is whether we would even trigger on such events. Remember the fundamental principle: "If it didn't trigger, it didn't happen." In other words, without a suitable trigger, the event will be lost forever and you might as well not have built the detector. Not good! In CMS a special trigger was implemented to search for energy deposits ($p_T > 50$ GeV) in the calorimeter that were present in the intervals between beam crossings (vetoing on signals from the beam position monitors on either side of the CMS detector) [95]. In a fill with 228 bunches per beam, 85 % of each orbit period (89 μs) was available for the search, falling to 16 % of the orbit period for a fill with 1,380 bunches.

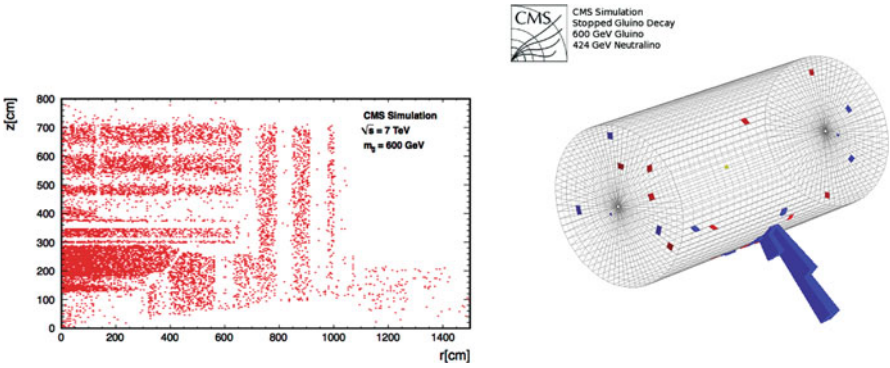

Fig. 24 CMS experiment: search for R-hadrons. *Left*: a map of the densest regions of the detector, where R-hadrons are most likely to stop. *Right*: a simulated R-hadron decay. Note that the pointing direction of the displayed calorimeter tower is not meaningful in the context of this search

Remarkably, it was possible to suppress backgrounds from sources such as beam-halo events, cosmic rays, and calorimeter noise to a very low level, around 1.5×10^{-6} Hz. Limits on various stopping particles are then obtained as a function of the particle masses and lifetimes [95].

7.3 Search for Microscopic Black Holes

The intriguing possibility of producing microscopic black holes at the LHC has attracted much attention, both in and outside the physics community. The production of black holes would be a signature of low-scale quantum gravity. There are many possible scenarios, leading to a small industry of models and accompanying simulation programs. The phenomenology of black-hole formation involves several subtleties, such as defining the fraction of the initial parton energy that is trapped within the event horizon, whether the black hole is rotating or not, whether there is a stable remnant, and so on. Black hole searches are based on signatures with rather broad interest, which involve events with very large total transverse energy and high particle multiplicity.

CMS has performed a black hole search [96] based on the kinematic variable S_T, which is the scalar sum of the transverse momenta of essentially all objects including E_T^{miss},

$$S_T = \sum_{i=j,\ \ell,\ \gamma,\ E_T^{\text{miss}}} p_T^i, \tag{13}$$

where j represents jets, ℓ represents isolated leptons, and γ represents isolated photons. Thresholds are applied to all objects. Distributions of S_T in 8 TeV data are

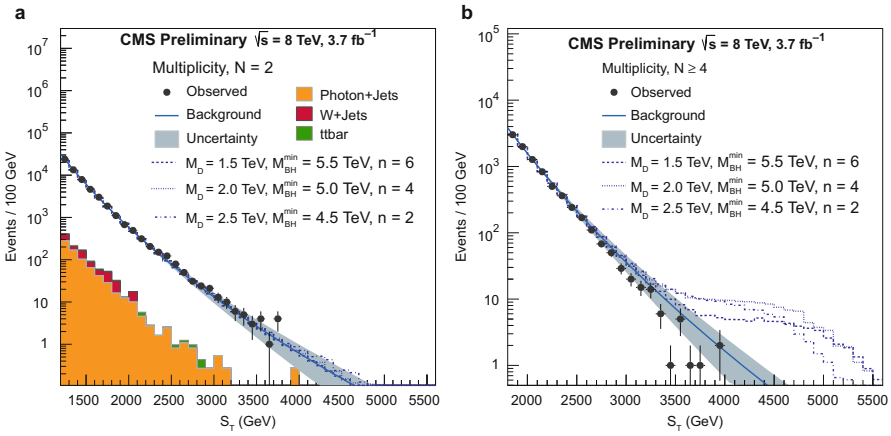

Fig. 25 Kinematic distributions from a CMS search for black holes [96]. The S_T variable is the scalar sum of the p_T values over essentially all objects (jets, isolated leptons, isolated photons, E_T^{miss}) in the event. The multiplicity N includes all objects except E_T^{miss}. *Left*: low-multiplicity ($N = 2$) control region. *Right*: example of a high-multiplicity ($N \geq 4$) signal region, with simulated black-hole signals

shown in Fig. 25; these distributions extend beyond 3 TeV. The background shape is obtained from a fit to low-multiplicity (denoted by N, where N does not include E_T^{miss}) events in data, with the restriction $1{,}200 < S_T < 2{,}800$ GeV. The shapes in the $N = 2$ and $N = 3$ samples are very similar, and a dedicated search for new physics in the $N = 2$ sample shows no signal. Figure 25 shows an example of a high-multiplicity sample, $N \geq 4$. The data are well described by the background shape, and black-hole signal shapes are included for reference. This study excludes black hole masses below 4–6 TeV, depending on the model.

8 Conclusions

With the LHC, we have an extraordinary tool for exploring the deep issues of electroweak unification and the Higgs sector, the mystery of the gauge hierarchy problem, and the nature of dark matter. In addressing these and other questions, we may (or may not) discover supersymmetry, extra dimensions, and new forces of nature. At a more basic level, the operations of the LHC at 7 and 8 TeV have been remarkably smooth. The upcoming run at 13 TeV promises to be one of the most important periods in the history of particle physics. There are no guarantees, but the potential for breakthroughs has never been greater. Your work and leadership will be critical in achieving these goals.

Acknowledgements I would like to express my deep thanks to all of the organizers and staff at the *69th Scottish Universities Summer School in Physics*. Their superb hospitality and hard work made the school a great success. The graduate students and postdocs at the school were outstanding, and I would like to thank them for their enthusiasm, advice, interest, and willingness to perform my skit, *Higgs in Hollywood*, at the banquet. I would also like to thank Robert Cahn and Jonathan Feng for their helpful advice. Finally, I would like to thank several of my UCSB colleagues and students for their many comments on the manuscript: Mark Srednicki, David Stuart, Manuel Franco Sevilla, Tom Danielson, Kristen Flowers, Adam Dishaw, and Jack Bradmiller-Feld.

References

1. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012)
2. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **716**, 30 (2012)
3. F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
4. P.W. Higgs, Phys. Lett. **12**, 132 (1964)
5. P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
6. G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
7. G. Aad et al., ATLAS Collaboration, JINST **3**, S08003 (2008)
8. S. Chatrchyan et al., CMS Collaboration, JINST **3**, S08004 (2008)
9. T.D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic Publishers, London, 1981), pp. 207–208
10. C. Amsler, T. DeGrand, B. Krusche, Quark model, in J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012)
11. H. E. Haber, Supersymmetry, part I (theory), in J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012)
12. P. Binétruy, *Supersymmetry: Theory, Experiment and Cosmology* (Oxford University Press, Oxford, 2006), p. 9
13. M. Papucci, J.T. Ruderman, A. Weiler, JHEP **1209**, 035 (2012)
14. G.F. Guidice, in *Perspectives on LHC Physics*, ed. by G. Kane, A. Pierce (World Scientific, Singapore, 2008), p. 155
15. J.L. Feng, Ann. Rev. Astron. Astrophys. **48**, 495 (2010)
16. G. Bertone, D. Hooper, J. Silk, Phys. Rep. **405**, 279 (2005)
17. D. Hooper, TASI2008 Lectures on Dark Matter, in *The Dawn of the LHC Era, Proceedings of the 2008 Theoretical Advanced Study Institute in Elementary Particle Physics*, Boulder, 2–27 June 2008, ed. by Tao Han (World Scientific, Singapore, 2009)
18. Y.A. Golfand, E.P. Likhtman, JETP Lett. **13**, 323 (1971)
19. D.V. Volkov, V.P. Akulov, Phys. Lett. B **46**, 109 (1973)
20. D.V. Volkov, V.A. Soroka, JETP Lett. **18**, 312 (1973)
21. J. Wess, B. Zumino, Nucl. Phys. B **70**, 39 (1974)
22. K.A. Olive, S. Rudaz, M.A. Shifman (eds.), *Proceedings of the International Symposium Celebrating 30 Years of Supersymmetry*, Minneapolis. Nuclear Physics, vol. B101 (Proc. Suppl.) (2001)
23. S.P. Martin, A Supersymmetry Primer. Adv.Ser.Direct.High Energy Phys. **21**, 1–153 (2010). doi:10.1142/9789814307505_0001, arXiv:hep-ph/9709356
24. L. Randall, R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999)
25. N. Arkani-Hamed, S. Dimopoulos, G.R. Dvali, Phys. Lett. B **429**, 263 (1998)
26. I. Aitchison, *Supersymmetry in Particle Physics* (Cambridge University Press, Cambridge, 2007)
27. H. Baer, X. Tata, *Weak Scale Supersymmetry* (Cambridge University Press, Cambridge, 2006)
28. M. Drees, R. Godbole, P. Roy, *Theory and Phenomenology of Sparticles* (World Scientific, Singapore, 2004)

29. J. Wess, J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, Princeton, 1992)
30. O. Buchmuller, P. de Jong, Supersymmetry, part II (experiment), in J. Beringer et al., Particle Data Group Collaboration, Phys. Rev. D **86**, 010001 (2012)
31. D. Green, *At the Leading Edge: The ATLAS and CMS LHC Experiments* (World Scientific, Singapore, 2010)
32. B. Niczyporuk, representing the Crystal Ball Collaboration, in *Proceedings of the 12th SLAC Summer Institute on Particle Physics*, Stanford, 1984, ed. by P. McDonough. SLAC Report No. 281: 513, 1985. See also Crystal Ball Collaboration, "Evidence for Narrow Massive State in the Radiative Decays of the Upsilon", SLAC-PUB-3380, 1984
33. B. Schwarzchild, *Physics Today*, p. 19, March 1985
34. G. Arnison et al., UA1 Collaboration, Phys. Lett. B **147**, 493 (1984)
35. S. Stepanyan et al., CLAS Collaboration, Phys. Rev. Lett. **91**, 252001 (2003)
36. K. Hicks, J. Phys. Conf. Ser. **9**, 183 (2005)
37. K.H. Hicks, Eur. Phys. J. H **37**, 1 (2012)
38. R.A. Schumacher, AIP Conf. Proc. **842**, 409 (2006)
39. A. Roodman, eConf C **030908**, TUIT001 (2003)
40. J.R. Klein, A. Roodman, Ann. Rev. Nucl. Part. Sci. **55**, 141 (2005)
41. G.L. Bayatian et al., CMS Collaboration, J. Phys. G **34**, 995 (2007)
42. S. Chatrchyan et al., CMS Collaboration, JHEP **10**, 132 (2011); Phys. Lett. B **701**, 535 (2011), CMS-PAS-EWK-10-012, CMS-PAS-EWK-11-010 and CMS-PAS-HIG-11-025
43. A. Czarnecki, J.G. Korner, J.H. Piclum, Phys. Rev. D **81**, 111503 (2010)
44. M. Jacob, G.C. Wick, Ann. Phys. **7**, 404 (1959)
45. J.D. Richman, Caltech Preprint CALT-68-1148 (1986)
46. S. Chatrchyan et al., CMS Collaboration, Eur. Phys. J. C **73**, 2404 (2013)
47. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **733**, 328 (2014)
48. Z. Bern et al., Phys. Rev. D **84**, 034008 (2011)
49. Z. Bern et al., Phys. Rev. D **87** 3, 034026 (2013)
50. G.L. Kane, C.F. Kolda, L. Roszkowski, J.D. Wells, Phys. Rev. D **49**, 6173 (1994)
51. A.H. Chamseddine, R.L. Arnowitt, P. Nath, Phys. Rev. Lett. **49**, 970 (1982)
52. J.A. Conley et al., Eur. Phys. J. C **71**, 1697 (2011)
53. J.A. Conley et al., arXiv:1103.1697 [hep-ph]
54. N. Arkani-Hamed et al., hep-ph/0703088 [HEP-PH]
55. J. Alwall, P. Schuster, N. Toro, Phys. Rev. D **79**, 075020 (2009)
56. J. Alwall, M.-P. Le, M. Lisanti, J.G. Wacker, Phys. Rev. D **79**, 015005 (2009)
57. D. Alves et al., LHC New Physics Working Group Collaboration, J. Phys. G **39**, 105005 (2012)
58. H.P. Nilles, Phys. Rep. **110**, 1 (1984)
59. P. Fayet, Phys. Lett. B **70**, 461 (1977)
60. H. Baer, M. Brhlik, C.-h. Chen, X. Tata, Phys. Rev. D **55**, 4463 (1997)
61. H. Baer, P.G. Mercadante, X. Tata, Y.-l. Wang, Phys. Rev. D **60**, 055001 (1999)
62. S. Dimopoulos, S.D. Thomas, J.D. Wells, Nucl. Phys. B **488**, 39 (1997)
63. J.R. Ellis, J.L. Lopez, D.V. Nanopoulos, Phys. Lett. B **394**, 354 (1997)
64. M. Dine, A.E. Nelson, Y. Nir, Y. Shirman, Phys. Rev. D **53**, 2658 (1996)
65. G.F. Giudice, R. Rattazzi, in *Perspectives on Supersymmetry* (World Scientific, Singapore, 1998), p. 355
66. P. Meade, N. Seiberg, D. Shih, Prog. Theor. Phys. Suppl. **177**, 143 (2009)
67. M. Buican, P. Meade, N. Seiberg, D. Shih, JHEP **0903**, 016 (2009)
68. G.R. Farrar, P. Fayet, Phys. Lett. B **76**, 575 (1978)
69. W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, I. Niessen, JHEP **1008**, 098 (2010)
70. W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, L. Motyka, I. Niessen, Int. J. Mod. Phys. A **26**, 2637 (2011)

71. M. Kramer, A. Kulesza, R. van der Leeuw, M. Mangano, S. Padhi, T. Plehn, X. Portell, Supersymmetry production cross sections in pp collisions at sqrt(s) = 7 TeV. 1326. arXiv:1206.2892 [hep-ph]
72. G. Aad et al., ATLAS Collaboration, Eur. Phys. J. C **72**, 1844 (2012)
73. C.G. Lester, D.J. Summers, Phys. Lett. B **463**, 99 (1999)
74. A. Barr, C. Lester, P. Stephens, J. Phys. G **29**, 2343 (2003)
75. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. Lett. **111**, 081802 (2013)
76. ATLAS Collaboration, Search for new phenomena in monojet plus missing transverse momentum final states using 10fb-1 of pp collisions at sqrts=8 TeV with the ATLAS detector at the LHC. ATLAS-CONF-2012-147
77. G. Aad et al., ATLAS Collaboration, Phys. Rev. Lett. **110**, 011802 (2013)
78. CMS Collaboration, Search for new physics in monojet events in pp collisions at sqrt(s)=8 TeV. CMS-PAS-EXO-12-048
79. B. Gomber, CMS Collaboration, Phys. Rev. Lett. **108**, 261803 (2012)
80. Y. Bai, P.J. Fox, R. Harnik, JHEP **1012**, 048 (2010)
81. A. Rajaraman, W. Shepherd, T.M.P. Tait, A.M. Wijangco, Phys. Rev. D **84**, 095013 (2011)
82. L. Randall, D. Tucker-Smith, Phys. Rev. Lett. **101**, 221803 (2008)
83. S. Chatrchyan et al., CMS Collaboration, EPJC **73**, 2568 (2013)
84. S. Chatrchyan et al., CMS Collaboration, Phys. Rev. Lett. **109**, 171803 (2012)
85. ATLAS Collaboration, ATLAS-CONF-2012-104
86. D.S.M. Alves, M.R. Buckley, P.J. Fox, J.D. Lykken, C.-T. Yu, Phys. Rev. D **87** 3, 035016 (2013)
87. ATLAS Collaboration, ATLAS-CONF-2013-024, ATLAS-CONF-2013-037, ATLAS-CONF-2012-171, and ATLAS-CONF-2012-167
88. S. Chatrchyan et al., CMS Collaboration, JHEP **1303**, 037 (2013). (Erratum-ibid. **1307**, 041, 2013)
89. ATLAS Collaboration, ATLAS-CONF-2012-105
90. H. Davoudiasl, J.L. Hewett, T.G. Rizzo, Phys. Rev. D **63**, 075004 (2001)
91. M.J. Strassler, K.M. Zurek, Phys. Lett. B **651**, 374 (2007)
92. T. Han, Z. Si, K.M. Zurek, M.J. Strassler, JHEP **0807**, 008 (2008)
93. N. Arkani-Hamed, S. Dimopoulos, JHEP **0506**, 073 (2005)
94. N. Arkani-Hamed, S. Dimopoulos, G.F. Giudice, A. Romanino, Nucl. Phys. B **709**, 3 (2005)
95. V. Khachatryan et al., CMS Collaboration, Phys. Rev. Lett. **106**, 011801 (2011)
96. S. Chatrchyan et al., CMS Collaboration, JHEP **1307**, 178 (2013)

Part III
Tools

Monte Carlo Event Generators

Michael H. Seymour and Marilyn Marx

Abstract Monte Carlo event generators are essential components of almost all experimental analyses and are also widely used by theorists and experiments to make predictions and preparations for future experiments. They are all too often used as "black boxes", without sufficient consideration of their component models or their reliability. In this set of three lectures we hope to open the box and explain the physical bases behind the models they use. We focus primarily on the general features of parton showers, hadronization and underlying event generation.

1 Motivation and Overview

Monte Carlo (MC) event generators are very widely used, especially by experimentalists in analyses but also by many theorists, who use them to make predictions for collider experiments and to develop techniques to propose to the experiments. MC are extremely important tools in High Energy Physics but unfortunately they are often used as "black boxes" whose outcome is treated as data. The aim of these lectures is to explain the physics behind event generators, which are mostly common between event generators but some differences will be highlighted.

As an example of the importance of MC, the majority of the recent Higgs discovery plots rely very strongly on MC predictions, to set limits on Higgses in certain parameter space regions as well as to discover them. This should be motivation enough to show that we need event generators for doing discovery as well as precision physics. Figure 1a shows the ATLAS diphoton invariant mass distribution consistent with a Standard Model Higgs boson of 126 GeV. One might ask if event generators are really still necessary when a distinct bump such as this one is visible. The answer is certainly yes, for example to quantify the significance of such a resonance and understand what particle it is. In the $H \to \gamma\gamma$ channel

M.H. Seymour (✉) • M. Marx
School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
e-mail: michael.seymour@manchester.ac.uk; lynn.marx@hep.manchester.ac.uk

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_8

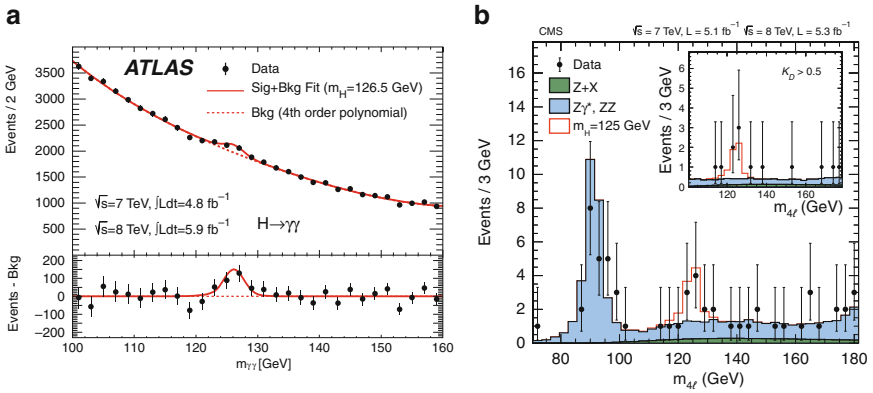

Fig. 1 Invariant mass distributions of (**a**) ATLAS $H \to \gamma\gamma$ and (**b**) CMS $H \to ZZ \to 4\ell$ candidates for the combined $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ data samples (Reproduced from [1, 2])

the resonance sits on a very steeply falling background where event generator predictions might be less important but all other discovery channels rely extremely heavily on event generators. Figure 1b shows the CMS four lepton invariant mass distribution from the "Golden Channel", where MC predictions are crucial for signal and background modelling.

The structure of a proton-proton collision at the Large Hadron Collider (LHC) as built up by event generators can be described by a few main steps. These are illustrated in Fig. 2 where two protons come in from either side and make a collision. The colour coding corresponds to the steps into which most event generators divide the process:

1. Hard process
2. Parton shower
3. Hadronization
4. Underlying event
5. Unstable particle decays

The first thing an experimentalist notices when studying proton-proton collisions is that most of them are "boring" in the sense that only a few soft hadrons are produced and most of the event goes out along the beam pipe direction. Only a tiny fraction of events contain a high momentum-transfer process of interest. It is therefore not feasible to simulate all possible proton-proton collisions but the simulation needs to be structured with a focus on deciding what hard process is wanted (a bit like triggers at experiments which decide which events to write to tape and which to discard).

This is done by starting the simulation at the heart of the collision and calculating from perturbation theory the probability distribution of a particular *hard scatter*, which is the highest momentum transfer process in the event. Simulating the hard

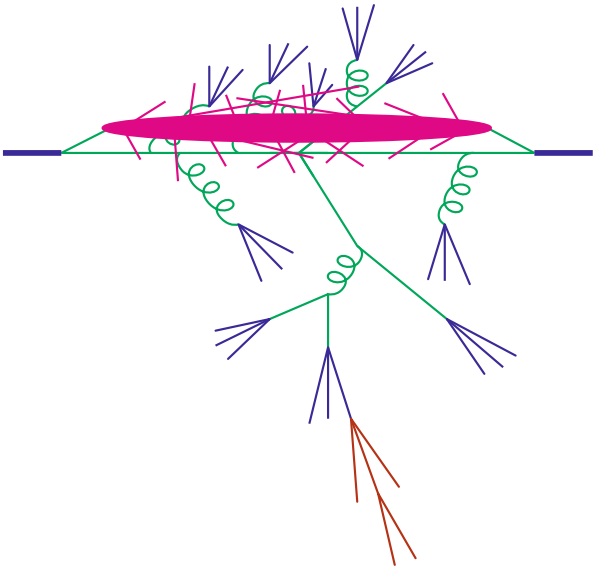

Fig. 2 Diagram showing the structure of a proton-proton collision, where the different shadings indicate the different stages involved in event generation

process is relatively straightforward because Parton Distribution Functions (PDFs) describe partons coming into the process and lowest order perturbation theory gives a probabilistic distribution of the outgoing partons.

A more interesting stage of event generation comes from asking what happens to the incoming and outgoing partons involved in the hard collision. This is described by the *parton shower* phase of event generators. The partons involved in the hard process are coloured particles, quarks and gluons. From Quantum Electrodynamics (QED) it is well known that scattered electric charges radiate photons, this is what is called Bremsstrahlung. In the same way, scattered colour charges radiate gluons and this happens for partons on their way in and out of a collision. The main difference to QED is that, due to the non-Abelian structure of $SU(3)$, gluons themselves are coloured and so an emitted gluon can itself trigger new radiation. This leads to an extended shower and the phase space fills up with (mostly) soft gluons. The parton shower can be simulated as a sequential step-by-step process that is formulated as an evolution in momentum transfer scale. The parton shower evolution starts from the hard process and works downwards to lower and lower momentum scales to a point where perturbation theory breaks down.

Here it is necessary to switch to *hadronization* models, which take account of the confinement of a system of partons into hadrons, which are seen in the detector. As well as the confinement of the produced partons, it is important to remember that the initial, uncoloured proton has had a coloured parton taken out if it and so it has been left in a coloured state. To get an idea of the space time structure

of a collision, consider the fact that in a proton's own rest frame it is a spherical bound state, but in the lab frame the two protons are moving towards each other with very high momentum and the Lorentz contraction flattens them into extremely thin pancakes. The collision happens at a point where these flat discs are completely overlapping each other in space time and so there is a very high probability that there will be other interactions apart from the hard interaction. This gives rise to the *underlying event*, which is made up of secondary interactions between proton remnants. It produces soft hadrons everywhere in the event, which overlie and contaminate the hard process that was already simulated.

The last component of event generation, which is usually not discussed in as much detail, is the fact that many of these hadrons are not stable particles but heavy resonances that then go on to decay. A lot of improvement has been made in the last 5 years to model these secondary particle decays.

Although some details differ, this brief overview of a process from hard collision to stable hadrons is effectively used by all current general purpose event generators, i.e. Herwig, Pythia and Sherpa. The lectures are organized into three main parts: parton shower, hadronization as well as underlying event and soft inclusive physics models. More details can be obtained from [3]. The classic textbook on the subject is [4].

2 Parton Showers

The basic idea of the parton shower is to set up in a probabilistic way a simulation of the cascade of partons that is produced by the colour charges that are accelerated in a scattering process, or created or annihilated in a pair creation process. The simulation of final state radiation (FSR) will be discussed, namely what happens to the partons as they leave the hard collision. Finally, it will be shown that the main ideas for FSR can also be applied to initial state radiation (ISR).

2.1 Divergence of QCD Emission Matrix Elements

First we want to look at the simplest, non-trivial Quantum Chromodynamics (QCD) process one can study, e^+e^- annihilation to jets. The tree-level cross section for e^+e^- annihilation to two partons ($q\bar{q}$) is finite and, from a QCD point of view, does not have much interest. However, the first correction to this process, namely e^+e^- to three partons is already very interesting. It is a good example of the more general statement that almost all QCD matrix elements diverge in different regions of phase space. It is the need to understand these divergences that will lead to the parton shower description of FSR. If we want to calculate the distribution of three partons in the final state, we need to sum two Feynman diagrams, shown in Fig. 3, at the amplitude level and then square them.

Fig. 3 Feynman diagrams for the process $e^+e^- \to q\bar{q}g$

One can calculate the differential cross section and write it, as shown in Eq. (1), in terms of the opening angle θ between the quark and the gluon, the energy fraction of the gluon $z_g = E_g/E_{g,max}$, the total $e^+e^- \to q\bar{q}$ cross section σ_0, the quark charge squared C_F and the QCD running coupling constant $\alpha_s \sim 0.1$,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta\,\mathrm{d}z_g} \sim \sigma_0\, C_F\, \frac{\alpha_s}{2\pi}\, \frac{2}{\sin^2\theta}\, \frac{1+(1-z_g)^2}{z_g}. \tag{1}$$

This formula has several interesting features. It has a factorized form as it is proportional to σ_0 and so one can think of this as a two step process: first the e^+e^- makes a two parton system, which in turn produces an extra gluon. Another important point is that it is not possible to calculate the whole probability as it is divergent in the collinear limit ($\theta \to 0, \pi$)[1] and in the soft limit ($z_g \to 0$).[2]

First we think about the physics of the collinear limit $\theta \to 0, \pi$ of QCD matrix elements. Parts of the previous equation can be separated into two pieces

$$\frac{2\,\mathrm{d}\cos\theta}{\sin^2\theta} = \frac{\mathrm{d}\cos\theta}{1-\cos\theta} + \frac{\mathrm{d}\cos\theta}{1+\cos\theta} \tag{2}$$

$$= \frac{\mathrm{d}\cos\theta}{1-\cos\theta} + \frac{\mathrm{d}\cos\bar{\theta}}{1-\cos\bar{\theta}} \tag{3}$$

$$\approx \frac{\mathrm{d}\theta^2}{\theta^2} + \frac{\mathrm{d}\bar{\theta}^2}{\bar{\theta}^2}, \tag{4}$$

where $\bar{\theta} = \pi - \theta$ is the angle between the antiquark and the gluon.[3] From the middle line, we can see that we have separated this into two independent terms, the first (second) term is only divergent in the $\theta \to 0$ ($\bar{\theta} \to 0$) quark (antiquark) limit.

[1] Assuming massless quarks; the massive case will be discussed later.

[2] It should be noted that here we parameterized the kinematics in the rest frame of the virtual photon, i.e. the rest frame of the total hadronic, three parton system. This might make it look like it is not Lorentz invariant, but one can show that the energy and angle dependence will always conspire in such a way that the final distributions are frame independent.

[3] This is an approximation that becomes exact in the limits of collinear or soft emission that we are interested in.

These terms can be approximated again to expose more clearly that this is a logarithmically divergent distribution. We have written this in a manner that appears sequential: in a first step the $q\bar{q}$ pair is produced and in a second step the gluon is radiated. The probability distribution of this gluon was separated out into the sum of two pieces, where one is associated to the quark direction and one to the antiquark direction. Rewriting the differential cross section, we can think of this as a system where each jet is evolving independently, each of which has a collinear factor:

$$\mathrm{d}\sigma = \sigma_0 \sum_{jets} C_F \, \frac{\alpha_s}{2\pi} \, \frac{\mathrm{d}\theta^2}{\theta^2} \, \mathrm{d}z \, \frac{1+(1-z)^2}{z}. \tag{5}$$

Here, we have set this up in terms of the opening angle θ as it is convenient, but we could build something of the exact same form that is proportional to θ^2, e.g. the transverse momentum of the gluon relative to the $q\bar{q}$ axis,

$$k_\perp^2 = z^2(1-z)^2\theta^2 E^2, \tag{6}$$

or the total invariant mass of the quark gluon system,

$$q^2 = z(1-z)\theta^2 E^2, \tag{7}$$

so that

$$\frac{\mathrm{d}\theta^2}{\theta^2} = \frac{\mathrm{d}k_\perp^2}{k_\perp^2} = \frac{\mathrm{d}q^2}{q^2}. \tag{8}$$

The choice of this variable is one of the important differences between various parton shower algorithms. In the limit of $\theta \to 0$, $k_\perp \to 0$, $q \to 0$, all these variables give the same leading approximation to the full cross section, Eq. (5), so in describing the cross section with leading accuracy they are equivalent and this is formally a free choice. However, each involves different sub-leading corrections to the leading approximation, so the choice can, and does in practice, have important consequences for the distributions produced.

2.2 Collinear Limit

In Eq. (5), the differential cross section was written in a factorized form, σ_0 times the sum over all hard partons that are involved in the process. One can show that this is a *universal* feature of QCD matrix elements and not unique to this e^+e^- case.

This differential cross section can be written in a universal way for an arbitrary hard process

$$\mathrm{d}\sigma = \sigma_0 \sum_{jets} \frac{\alpha_s}{2\pi} \, \frac{\mathrm{d}\theta^2}{\theta^2} \, \mathrm{d}z \, P(z,\phi) \, \frac{\mathrm{d}\phi}{2\pi}, \tag{9}$$

where z is the energy fraction of the parton, ϕ is the azimuthal angle of the splitting around the axis of the parent parton and $P(z, \phi)$ is known as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting kernel, which depends on flavour and spin. To give a few spin-averaged (no ϕ dependence) examples of the latter,

$$P_{q\to qg}(z) = C_F \frac{1+z^2}{1-z}, \tag{10}$$

$$P_{q\to gq}(z) = C_F \frac{1+(1-z)^2}{z}, \tag{11}$$

$$P_{g\to gg}(z) = C_A \frac{z^4+1+(1-z)^4}{z(1-z)}, \tag{12}$$

$$P_{g\to q\bar{q}}(z) = T_R \left(z^2+(1-z)^2\right). \tag{13}$$

To use the collinear limit, we do not have to take the partons to be exactly collinear, it is sufficient that the opening angle is much smaller than any angles involved in the hard process. In this limit we get this universal behaviour.

However, we still cannot describe this process probabilistically because Eq. (9) still diverges when integrated over all possible angles. To understand the physics behind this divergence, let us take a step back and think about what we mean by having a parton in the final state of our process. As we will discuss later, partons produce jets with a finite spread of hadrons. The hadronic state produced by two exactly collinear partons is identical to that produced by a single parton with their total momentum (and colour). Without yet going into the details of the hadronization process, let us assume that there is some value of momentum of one parton transverse to the axis defined by another, below which they cannot be resolved[4] and only calculate the probability distribution for *resolvable* partons. That is, we introduce an arbitrary parameter, Q_0, which describes whether two partons are resolvable from each other or not. If $k_\perp > Q_0$, we call them resolvable and use perturbation theory to describe them as two separate particles. If $k_\perp < Q_0$ we say that they are indistinguishable from a single parton with the same total momentum.

Now we can calculate the total probability for resolvable emission, which must be finite. Since unresolvable emission is indistinguishable from non-emission, we must add together the virtual (loop) correction to the original process and the integral of the emission probability over the unresolved region. Each is divergent, but the virtual divergence is negative and exactly cancels the real divergence. So although each of them is divergent, their sum is finite and obeys unitarity:

$$P(resolved) + P(unresolved) = 1. \tag{14}$$

[4]Note that we are not talking about our experimental ability to resolve jets, but an *in principle* indistinguishability of exactly collinear partons.

It is important to note that this fact is derived from Quantum Field Theory (QFT) and is not just assumed. One can encode the non-emission probability by something called a Sudakov form factor (SFF), which is a key ingredient of the MC. This SFF represents the probability that a given parton does not radiate resolvable gluons and has an exponential form. The probability of emission between q^2 and $q^2 + \mathrm{d}q^2$ is

$$\mathrm{d}\mathscr{P} = \frac{\alpha_s}{2\pi}\,\frac{\mathrm{d}q^2}{q^2}\int_{\frac{Q_0^2}{q^2}}^{1-\frac{Q_0^2}{q^2}} \mathrm{d}z\, P(z) \equiv \frac{\mathrm{d}q^2}{q^2}\,\bar{P}(q^2), \tag{15}$$

and the probability of no emission between Q^2 and q^2 is defined to be $\Delta(Q^2, q^2)$. This gives the evolution equation

$$\frac{\mathrm{d}\Delta(Q^2, q^2)}{\mathrm{d}q^2} = \Delta(Q^2, q^2)\,\frac{\mathrm{d}\mathscr{P}}{\mathrm{d}q^2}, \tag{16}$$

$$\Rightarrow \Delta(Q^2, q^2) = \exp\left(-\int_{q^2}^{Q^2}\frac{\mathrm{d}k^2}{k^2}\,\bar{P}(k^2)\right). \tag{17}$$

This has a very similar form to the well-known formula of radioactive decay, where an atom has a (constant) probability λ per unit time to decay:

$$P(\text{no decay after time } T) = \exp\left(-\int^{T}\mathrm{d}t\,\lambda\right). \tag{18}$$

The Sudakov form factor, $\Delta(Q^2, Q_0^2) \equiv \Delta(Q^2)$, represents the probability of emitting no resolvable radiation at all:

$$\Delta_q(Q^2) \sim \exp\left(-C_F\,\frac{\alpha_s}{2\pi}\log^2\frac{Q^2}{Q_0^2}\right), \tag{19}$$

which becomes very small for large Q^2, reflecting the fact that a quark formed at a high scale is extremely unlikely to be unaccompanied by *any* gluons.

2.3 Multiple Emission and Running Coupling

We can use the universality of the DGLAP splitting function to calculate the probability that, given an initial emission, a certain gluon radiates another gluon that is more collinear than the first one. This way we can attach more and more partons, as is shown in an example in Fig. 4. We can take the different building blocks that we have just derived, namely the tree-level splitting function and the SFF, which tells us the non-emission probability, and use them to construct the probability distribution

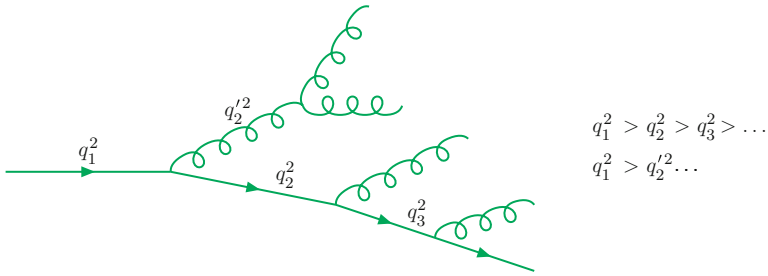

Fig. 4 Diagram showing multiple gluon emission off an initial quark line

of any complicated final state. It should be noted that one important point that needs to be specified is the initial condition. It tells us how large the initial value q_1 of the evolution variable can be. This is the only process dependent factor in the parton shower and we will come back to this later.[5]

To quickly touch on higher order loop corrections to emitted gluons, one can absorb a tower of higher order logarithmic corrections by replacing α_s by $\alpha_s(k_\perp^2)$. This is because at each higher order, α_s^n, one encounters terms like $\beta_0^n \ln^n(k_\perp^2/\mu^2)$, where β_0 is the leading coefficient of the QCD beta-function and μ is the renormalization scale at which α_s is defined. If μ^2 is very different to $k_\perp^2$ these terms are very large, spoiling the convergence of perturbation theory. But if the emission vertex is evaluated with α_s replaced by $\alpha_s(k_\perp^2)$, these terms are absorbed and effectively resummed to all orders. Thus, taking account of the increasing coupling at small $k_\perp$, the parton evolution is expected to fill the phase space with soft (low transverse momentum) gluons as it becomes increasingly "easy" (i.e. the probability becomes high) to emit very soft gluons. This further means that the Q_0 parameter is a very important physical parameter that constrains the parton shower algorithm. In order to use perturbation theory, one has to ensure that Q_0 is much larger than Λ_{QCD}, which is $\sim\mathcal{O}(200\,\text{MeV})$, and so it should be of order 1 or a few GeV.

2.4 Soft Limit and Angular Ordering

Now we want to move from the collinear limit to the soft limit, which is the other limit in which QCD matrix elements diverge. There is also a factorisation theorem for the soft limit but it has a very different form in the sense that it is universal only at the amplitude level, not at the cross section level.

[5]One can think of this initial condition as anything that parametrizes the "collinear-ness" of an emission process, e.g. the virtuality, how far off-shell this particle was. If we do not know how a coloured particle was produced, we cannot know how it radiates.

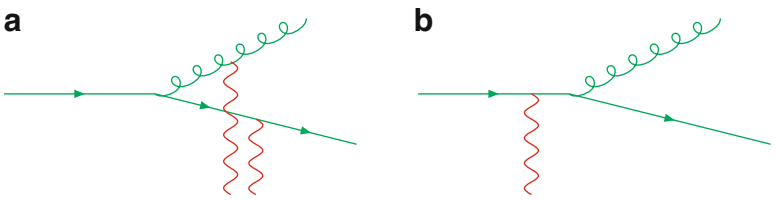

Fig. 5 Diagram describing the soft limit: the soft gluon may be emitted by either of the outgoing partons, at the amplitude level (**a**), but the sum of the two diagrams is as if it was emitted by a single parton with the same total momentum and colour (**b**)

Consider a quark coming out of a hard process that radiates a hard gluon. We want to know what the distribution of soft gluons radiated from this system looks like. Going back to QFT, there are two possible Feynman diagrams that contribute, which we illustrate by the single diagram in Fig. 5a. The soft gluon is attached to the hard gluon in one diagram and to the quark in the other. In either of these cases there is a factorisation theorem telling us that the amplitude for that process can be written as the amplitude to produce the hard gluon times a universal factor describing the radiation of the soft gluon. However, unlike in the collinear limit, we need to sum these two diagrams before we square them and since the amplitudes that they represent have similar magnitudes there will be quantum interference between them. At first this was thought to spoil the picture of independent evolution, as described before in the collinear limit. Actually this is not the case because the radiation from these two is coherent. If we sum up the two diagrams at the amplitude level and square them, the radiation pattern from this pair of partons is identical at large angles to the radiation pattern from a single quark with the same total colour charge and same total momentum as the pair of partons would have had if they were on-shell as shown on Fig. 5b. So at large angles, the gluons essentially only see the total colour charge, they cannot resolve the colour charges of the individual partons. On the other hand, when the opening angle to one of the hard partons is small, the corresponding single diagram dominates. We can incorporate this into our collinear parton shower algorithm by ordering in the opening angle, which will therefore correctly describe the soft limit.

We can conclude that it is possible to construct a parton shower that can describe correctly both the collinear and the soft limits of QCD matrix elements by using the opening angle as the evolution variable and describing a wide angle gluon as if it was emitted before the internal structure of the jet has built up.[6]

[6]At this point, we assume that the quarks are massless. We will see that the picture does not change radically for massive quarks, although for quarks heavier than Λ_{QCD} the quark mass actually plays a similar role to the resolution scale that we have discussed, cutting off collinear emission.

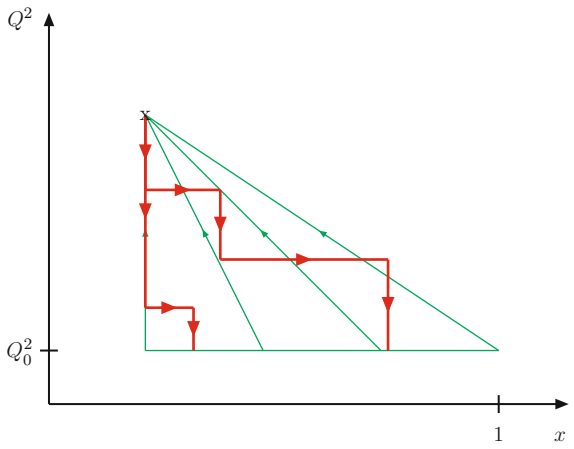

Fig. 6 The *green lines* illustrate the flow of information in analytic solutions of the DGLAP evolution equation, which yields the value of the parton distribution function at a given value of x and Q^2 as a function of its values at some lower value of Q^2, Q_0^2, and all higher values of x. The *red lines* illustrate typical backward evolution paths that lead to the same x and Q^2 value: each path corresponds to one event and each corner on the path to one emitted parton

2.5 Initial State Radiation

FSR is fully inclusive in the sense that we want to generate the distribution of all possible parton radiation. For ISR the goal is different. Here, we want to be able to choose the hard process and ask what radiation this process is accompanied by. So even though the physics involved in ISR and FSR is essentially the same, we have certain kinematic constraints for ISR, e.g. we know x and Q^2, and we therefore do not want to generate all possible distributions but only those subject to having a fixed parton momentum at the end of the process. As illustrated in Fig. 6, we can reformulate the evolution as a *backward evolution*, which probabilistically undoes the DGLAP evolution equations. We start from a particular x and Q^2 point and work down in q^2 and up in x towards the incoming hadron, asking progressively, what is the probability distribution of radiation that accompanies a parton of this flavour and kinematics. In the end, one finds that this algorithm is identical to FSR, but with $\Delta_i(Q^2, q^2)$ replaced by $\Delta_i(Q^2, q^2)/f_i(x, q^2)$.

2.6 Hard Scattering and Colour Coherence

We need to set the initial conditions for parton showers and here the colour coherence that we already talked about when studying the soft limit of QCD matrix elements, is important too. We take the example of quark-antiquark pair

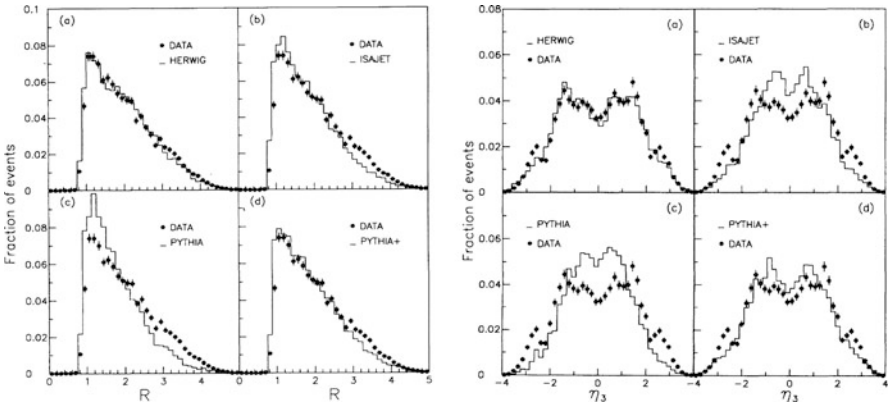

Fig. 7 Distributions from the CDF analysis demonstrating colour coherence in three-jet events. R is the separation, in $\eta - \phi$ space, between the second and third hardest jets and η_3 is the pseudorapidity of the third hardest jet (Reproduced from [5])

annihilation, say $u\bar{u} \to d\bar{d}$. The rules of perturbative QCD tell us that the quarks are in the fundamental representation of $SU(N_C = 3)$ and the gluons are in the adjoint representation, which has $N_C^2 - 1$ (= 8 for $SU(3)$) colours. Up to corrections of order $1/N_C^2$, we can think of a gluon as carrying a fundamental colour and a fundamental anti-colour label. In that approximation, the colour structure of this $q\bar{q}$ annihilation process looks completely different to the flavour structure (the "upness" is annihilated and becomes "downness", but the "redness" of an incoming quark gets transferred onto the s-channel gluon and thence onto the outgoing quark). The effect of the colour structure is best illustrated by contrasting a $u\bar{u} \to d\bar{d}$ event in which the d quark goes in the forward direction of the u quark with a $u\bar{d} \to u\bar{d}$ event with the same kinematics. In $u\bar{u} \to d\bar{d}$, the colour of the quark has only been scattered through a small angle and any emission from it is confined to angles smaller than the scattering angle. In $u\bar{d} \to u\bar{d}$, the colours in the initial state annihilate each other and a new colour-anticolour pair is created. Emission from both lines fills the whole of phase space. In general, the colour line of any parton can be traced through the hard process to find the parton to which it is colour connected: its colour partner. Emission from each parton is confined to a cone stretching to its colour partner and the colour coherence limits parton radiation in certain regions of phase space.

The CDF 3-jet analysis [5] is a classic experimental example that demonstrates that colour coherence is a real effect. In this analysis, they required two hard jets with $p_{t,1} > 110\,\text{GeV}$ but only $p_{t,3} > 10\,\text{GeV}$. To map out the kinematics of this third jet, they looked at the pseudorapidity and jet-separation distributions, which can be seen in Fig. 7. In these plots,[7] Herwig has colour coherence built in and predicts a dip between the hardest and second hardest jet, which was also

[7]The MC predictions here have been put through detector simulation.

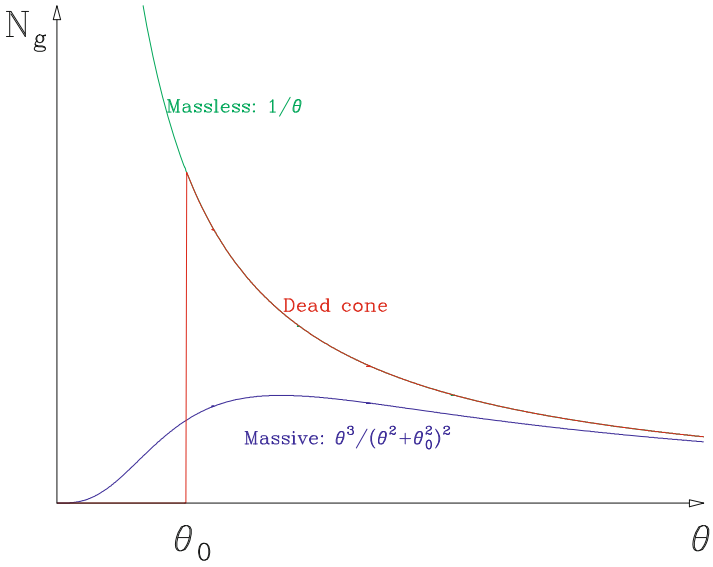

Fig. 8 Number of radiated gluons, N_g, as a function of the opening angle θ for massless (*green*) and massive (*blue*) quarks

seen by the data. The other two MCs shown, Pythia and Isajet, did not have this colour coherence built in. In this case, radiation was allowed to go everywhere and for kinematical reasons the radiation actually prefers to go in the central region. This misprediction prompted the Pythia authors to provide a better model, which was then called Pythia+ and had a partial treatment of colour coherence added. This feature is nowadays part of Pythia by default.

2.7 Heavy Quarks

So far we have only talked about the case where quarks are lighter than the confinement scale, for which their mass does not play an important role in their evolution because it is the confinement itself that provides a cut-off. But if we want to calculate the emission pattern from heavy quarks there is another colour coherence effect at work. Figure 8 shows the emission pattern of a final state quark. The massless case, drawn in green, goes as $1/\theta$ and diverges as we go to smaller angles. The pattern from a massive quark with the same momentum, drawn in blue, becomes similar at large angle (this is again the colour coherence effect – wide angle gluons only see the total colour) but as we get closer to $\theta_0 = \frac{m_q}{E_q}$ there is a smooth suppression and the true emission pattern turns over and goes to 0 at small angles. This means that a massive quark does not radiate at all in its forward direction. This blue curve is what is implemented in most current event generators but sometimes

people still talk about the "dead cone effect", which was the implementation in earlier generators. It is the somewhat brutal approximation of treating the massive quark as massless up to a certain angle and then just turning off radiation in the forward direction completely. This can be shown to give the correct total amount of radiation, but can be seen to be a very crude approximation for the distribution of that radiation.

More often a quasi-collinear splitting is used [6, 7], which has a smooth suppression in the forward region,

$$\mathrm{d}\mathscr{P}_{\widetilde{ij}\to ij} = \frac{\alpha_s}{2\pi}\,\frac{\mathrm{d}\tilde{q}^2}{\tilde{q}^2}\,\mathrm{d}z\,P_{\widetilde{ij}\to ij}(z,\tilde{q}^2). \tag{20}$$

For reference, we give the splitting functions for massive quarks and spartons[8]:

$$P_{q\to qg}(z,\tilde{q}^2) = \frac{C_F}{1-z}\left[1+z^2-\frac{2m_q^2}{z\tilde{q}^2}\right], \tag{21}$$

$$P_{g\to q\bar{q}}(z,\tilde{q}^2) = T_R\left[1-2z(1-z)+\frac{2m_q^2}{z(1-z)\tilde{q}^2}\right], \tag{22}$$

$$P_{\tilde{g}\to\tilde{g}g}(z,\tilde{q}^2) = \frac{C_A}{1-z}\left[1+z^2-\frac{2m_{\tilde{g}}^2}{z\tilde{q}^2}\right], \tag{23}$$

$$P_{\tilde{q}\to\tilde{q}g}(z,\tilde{q}^2) = \frac{2C_F}{1-z}\left[z-\frac{m_{\tilde{q}}}{z\tilde{q}^2}\right]. \tag{24}$$

2.8 Colour Dipole Model

So far we have been talking about "conventional" parton showers where you start from the collinear limit of QCD matrix elements and modify it to incorporate soft gluon coherence. In the colour dipole model (CDM) the starting point is somewhat different as it tries to understand the soft radiation first and to then modify that in such a way that you also get the collinear limit right. The idea is to start from the large N_C approximation where a gluon is treated as a colour-anticolour pair or dipole. The emission of soft gluons from such a dipole is universal (and classical)

$$\mathrm{d}\sigma \approx \sigma_0\,C_A\,\frac{\alpha_s(k_\perp)}{2\pi}\,\frac{\mathrm{d}k_\perp^2}{k_\perp^2}\,\mathrm{d}y, \tag{25}$$

[8] Since the gluon is massless $P_{g\to gg}$ is unchanged from Eq. (12).

where $y = -\log\tan\frac{\theta}{2}$ is the rapidity. In this model, we think of colour-anticolour *pairs* as radiating, so the $q\bar{q}$ pair together radiate a gluon. This creates an additional colour line, separating the system into a qg dipole and a $g\bar{q}$ dipole, which go on to radiate further. This way subsequent dipoles continue to cascade and instead of a $1 \to 2$ parton splitting, like in parton showers, you have a $1 \to 2$ dipole splitting, which corresponds to a $2 \to 3$ parton splitting.

One different feature of the CDM is that there is no explicit ISR. The hadron remnant forms a colour dipole with the scattered quark, which is treated like any other dipole, except for the fact that the remnant is an extended object. This radiates gluons, but since it is an extended object it does not radiate in its forward direction. The radiation looks like FSR from the outgoing remnant rather than ISR from the ingoing quark, but one can show that they are equivalent, with the suppression of radiation from the remnant interpreted as the suppression due to parton distribution function effects in ISR.

Most of the parton shower implementations that have appeared in the last few years [8–12] are based on this dipole cascade picture [13, 14].

2.9 Matrix Element Matching

The parton shower method is an approximation derived from QCD that is valid in the collinear and soft limits. It describes the bulk of radiation well but very often one uses event generators to search for new physics, to predict backgrounds or to model features of the signal processes and do precision physics, like e.g. the top mass measurement, measuring multi-jet cross sections etc. In these applications, you are not only interested in very soft and collinear emission but in systems of hard, well-separated jets. Therefore many of the applications of parton shower event generators are pushing them into regions of phase space where they are least reliable, i.e. away from the soft and collinear approximations and more into regions where fixed order matrix elements should describe those processes better. In order to improve their predictions, one would like to get simultaneously (at least) next-to-leading order (NLO) normalization, a good description of hard multi-jet systems but also match that with a good parton shower of the internal structure of those jets – i.e. the best of all worlds. Achieving this is known as matrix element matching and is one of the areas where MCs have developed the most in the last 5 years. Several methods have been proposed to combine tree-level matrix elements for several jet multiplicities simultaneously, with parton showers to describe the internal structure of the jets and the pattern of soft radiation between the jets (the "intrajet" and "interjet" event structure respectively) without double counting (the buzz-words are CKKW [15,16] or CKKWL [17] and MLM [18] matching). Alternatively, two methods have been proposed to combine lowest-multiplicity NLO matrix elements with parton showers, again without double counting (MC@NLO [19] and POWHEG [20]). The current state of the art is progress towards NLO multi-jet matching [8, 21–23], which is needed for many applications at the LHC.

2.10 Summary of Available Programs

We briefly mention some of the parton shower-related features of the different MC programs that are available.

Older programs that are still sometimes seen but not supported any more:

- Pythia 6.2 [24]: traditional q^2 ordering, veto of non-ordered final state emission, partial implementation of angular ordering in initial state, big range of hard processes.
- HERWIG [25]: complete implementation of colour coherence, NLO evolution for large x, smaller range of hard processes.
- Ariadne [26]: complete implementation of colour-dipole model, best fit to HERA data, interfaced to Pythia for hard processes.

Supported and new programs:

- Pythia 6.3 [27]: p_t-ordered parton showers, interleaved with multi-parton interactions, dipole-style recoil, matrix element for first emission in many processes.
- Pythia 8 [28]: new program with many of the same features as Pythia 6.3, many 'obsolete' features removed.
- Sherpa [29]: new program built from scratch: p_t-ordered dipole showers, multi-jet matching scheme (CKKW) to AMAGIC++ built in.
- Herwig++ [30]: new program with similar parton shower to HERWIG (angular ordered) plus quasi-collinear limit and recoil strategy based on colour flow, spin correlations.

In addition, dipole showers are available as optional plug-ins to both Herwig++ [31] and Pythia [12].

2.11 Summary

The basic idea of parton showers is very simple: accelerated colour charges radiate gluons, but since the gluons themselves are also charged we get an extended cascade developing. This cascade is modeled as an evolution downward in momentum scale. As we approach the non-perturbative limit, we get more and more radiation and the phase space fills with soft gluons. The probabilistic language is derived from factorization theorems of the full gauge theory. Colour coherence is a fact of life: do not trust those who ignore it!

Modern parton shower models are very sophisticated implementations of perturbative QCD, but they would be useless without the hadronization model, which will be discussed next.

3 Hadronization

Everything we have studied so far was based on perturbative QCD but partons are not the final state particles that come out of the collision as they cannot propagate freely. We know that hadrons are the physical final state particles, but we do not know how to calculate them, so we need a model to describe how partons are confined into hadrons – this is called hadronization. The two main models in use are the String Model, implemented in Pythia, and the Cluster Model, implemented in Herwig and Sherpa. These models will be described in more detail later but we will first look at some of the physics behind hadronization models and how they have developed.

3.1 Phenomenological Models

We start with an experimental observation, namely e^+e^- annihilation to two jets, which is the majority of hadronic e^+e^- events, and study the distribution of hadrons with respect to the axis formed by the two jets. You can measure the rapidity, y, and the transverse momentum, p_t, of hadrons relative to that axis. If you plot the number of hadrons as a function of y, sketched on the left in Fig. 9, you find that it is roughly flat up to some maximum value and then falls off very quickly. However looking at the right side of Fig. 9 where the number of hadrons is sketched as a function of p_t, it can be seen that this distribution is roughly Gaussian with a narrow width of 1 or 2 GeV. This means that most hadrons are produced with very low p_t.

One can make a very simple model based on this observation and estimate the hadronization correction to perturbative quantities. The energy of the jet is

$$E = \int_0^Y \mathrm{d}y\, \mathrm{d}^2 p_t\, \rho(p_t^2) \cosh y = \lambda \sinh Y, \tag{26}$$

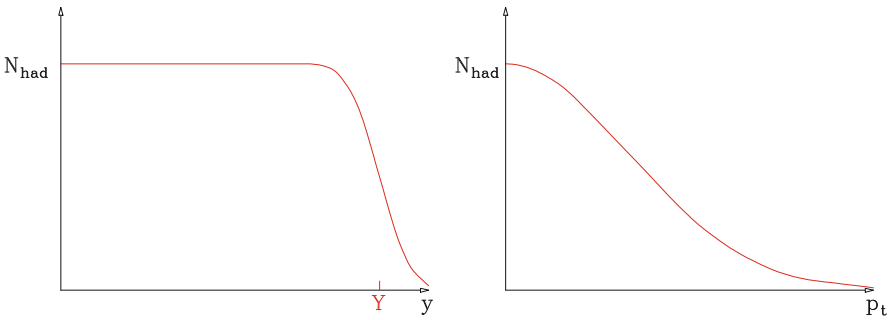

Fig. 9 The number of hadrons, N_{had}, sketched as a function of rapidity, y, and transverse momentum, p_t

where Y is the maximum rapidity of hadrons in the jet and λ is their mean transverse momentum, given by

$$\lambda = \int \mathrm{d}^2 p_t \, \rho(p_t^2) \, p_t, \tag{27}$$

which can be estimated from Fermi motion where $\lambda \sim 1/R_{had} \sim m_{had}$. The longitudinal momentum can be calculated in the same way:

$$P = \int_0^Y \mathrm{d}y \, \mathrm{d}^2 p_t \, \rho(p_t^2) \, \sinh y = \lambda(\cosh Y - 1) \sim E - \lambda. \tag{28}$$

The jet acquires a non-perturbative mass, given by

$$M^2 = E^2 - P^2 \sim 2\lambda E, \tag{29}$$

from which it can be seen that the non-perturbative invariant jet mass is proportional to the square root of its energy. This non-perturbative component is an important contribution, e.g. a 10 GeV contribution for 100 GeV jets. Since these corrections are so important, we need a precise model to predict them.

The first set of models that were developed to describe hadronization were the so-called Independent Fragmentation Models (or "Feynman-Field" models) and they are a direct implementation of the procedure described above. The longitudinal momentum distribution is an arbitrary fragmentation function, a parametrization of data. The transverse momentum distribution is assumed Gaussian and the model just recursively applies $q \to q' + had$ and hooks up the remaining soft q and $\bar{q}$ until the whole jet is hadronized. This model can describe $e^+e^- \to 2$ jet events by construction, but has a lot of disadvantages: it is strongly frame dependent, there is no obvious relation with perturbative emission, it is not infrared safe and it is not a model for confinement.

3.2 Confinement

We know that in QCD we have asymptotic freedom: at very short distances a $q\bar{q}$ pair becomes more and more QED-like, but at long distances the non-Abelian gluon self-interaction makes the field lines attract each other, as sketched in Fig. 10. As two colour charges are pulled apart the field lines do not spread out and we get a constant force or a linearly rising potential. One would have to invest an infinite amount of work to pull them apart – this is the signal of confinement.

This interquark potential (or string tension) can for example be measured from quarkonia spectra, as shown in Fig. 11, or from lattice QCD. The string tension κ is found to be roughly 1 GeV/fm.

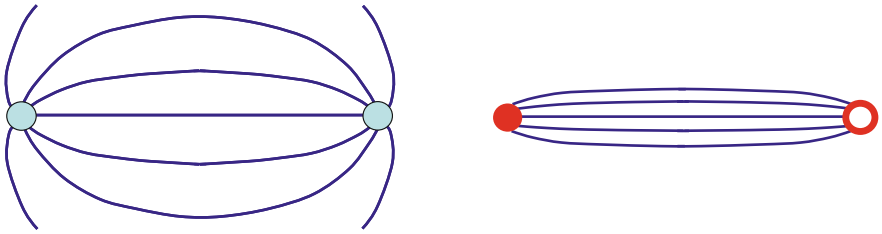

Fig. 10 Field lines in QED (*left*) and QCD (*right*) between a charge and an anti-charge

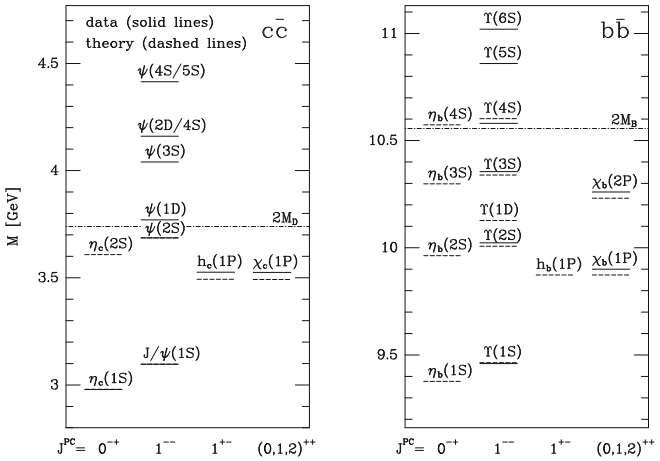

Fig. 11 Charmonium and bottomonium spectra, the nS ($n = 1, 2, \ldots$) energy levels are roughly equally spaced (in more detail, they go like $\sim n^{2/3}$) (Reproduced from [4])

3.3 String Model

The first step in understanding the structure of hadrons is to take this string picture very literally and to look at the space-time diagram of a meson ($q\bar{q}$), outlined in Fig. 12.

At some point in time the quark and the anti-quark (both assumed massless) are at the same point in space and they are flying apart at the speed of light, hence up a 45° diagonal in the space-time diagram. As they grow further apart they lay out a string between them that has a constant tension until the potential stored in this string uses up all their kinetic energy. At this point they turn around and the potential energy in the string accelerates the quarks towards each other until they meet back at the starting point, pass through each other, and the whole process starts over. These are the so-called "yo-yo" modes.

It is a nice exercise in Lorentz transformations to think about what this process looks like when you are not in the rest frame but in a boosted frame, e.g. when the meson is moving to the right. The two points at which the quarks' directions

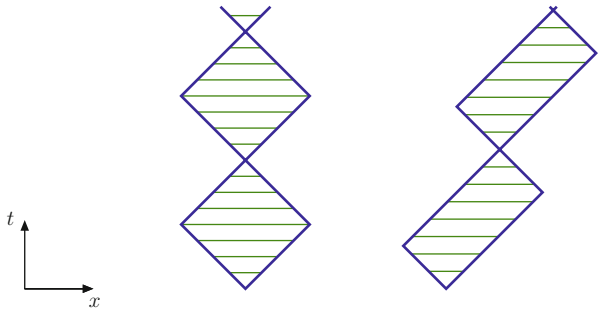

Fig. 12 Cartoon of string model, quark and antiquark moving apart from each other as seen in the meson rest frame (*left*) and in a boosted frame (*right*)

reverse are not simultaneous and the string spends part of its cycle simply moving to the right, without expanding or contracting, and transferring momentum from one quark to the other. You can measure the speed of the meson from the slope of the line of its centre of mass. The area of these squares and rectangles is the same in both frames, i.e. it is Lorentz invariant, and obeys the area law

$$m^2 = 2\kappa^2 \text{ area}. \tag{30}$$

The Lund String Model uses this picture as a model of hadronization. In the original, simple version of the model we start by ignoring gluon radiation. e^+e^- annihilation is then a point-like source of $q\bar{q}$ pairs. In principle this system also has yo-yo modes, but in practice the space-time volume swept out is so large that another effect is able to dominate: in the intense chromomagnetic field of the string between the $q\bar{q}$ pair it is possible for additional $q\bar{q}$ pairs to spontaneously tunnel out of the vacuum. The chromomagnetic field is strong enough that the quarks are accelerated away from each other before they have time to re-annihilate.[9] The effect is that the string separates into two strings: it breaks. By analogy with a similar process in QED you can estimate the probability of this happening, as

$$\frac{\mathrm{d}(\text{Probability})}{\mathrm{d}x\,\mathrm{d}t} \propto \exp\left(\frac{-\pi m_q^2}{\kappa}\right). \tag{31}$$

The mass dependence of this equation means that, for example, strange quarks will tunnel out less often than light quarks.

When we have a lot of energy available it is likely that we will produce many hadrons, as the expanding string breaks into mesons long before the yo-yo point. Thus the original $q\bar{q}$ system has fragmented into a system of hadrons – this is the

[9] This is somewhat analogous to Hawking radiation.

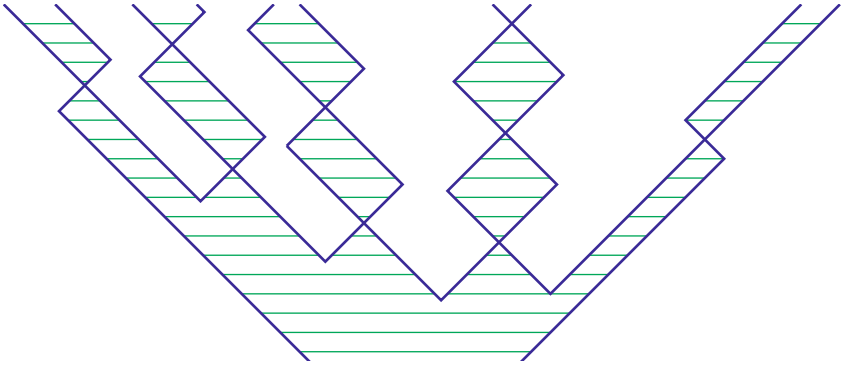

Fig. 13 Cartoon of the string corresponding to a high energy $q\bar{q}$ event breaking up into hadrons

basic ingredient of the Lund String model illustrated in Fig. 13. The space time structure of this breakup is very interesting, as the breaks are causally disconnected, so they don't know about each other and there can be no causal correlations between them. The Lorentz invariance and acausality give strong constraints on this hadronization process. In the end, we get a fragmentation function for hadrons with a constrained form with two adjustable parameters, a_α and a_β,

$$f(z) \propto z^{a_\alpha - a_\beta - 1}(1-z)^{a_\beta}. \tag{32}$$

The tunnelling probability then becomes

$$\exp\left(-b(m_q^2 + p_t^2)\right), \tag{33}$$

where the main tuneable parameters of the model are a_α and a_β, as described above, b, customarily called the "Lund b parameter" related to the string tension (which can be seen to control the width of the p_t distribution) and m_q^2, the masses of the individual quarks.

An important new feature of this model, relative to independent fragmentation, is that it is universal and predictive: having been tuned to describe quark events it can predict the confinement of gluons. This is again related to the colour structure. In a three parton ($q\bar{q}g$) system, the quark is colour-connected to the anticolour index of the gluon and the colour index of the gluon is connected to the antiquark. Thus the gluon makes a corner, or "kink", on the string. The acausality means that the breakup of the string is universal, but the Lorentz boost of a string means that the hadrons it produces go preferentially in the direction of its motion. Therefore most hadrons that the first string segment produces will go between the quark and the gluon, most hadrons from the second will go between the gluon and the antiquark and only very few hadrons will go between the quark and the antiquark.

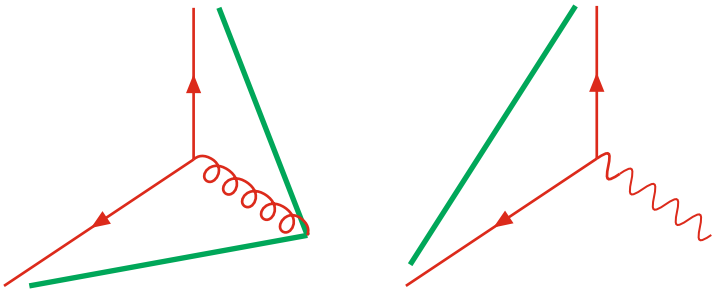

Fig. 14 String structure for $q\bar{q}g$ (*left*) and $q\bar{q}\gamma$ (*right*) events

This definite prediction of the string model is known as the *string effect* and can be seen experimentally, e.g. at the PETRA and LEP experiments, by comparing 3-jet events to 2-jet + photon events, which can be represented as in Fig. 14, where hadrons prefer to be between the quark and the antiquark.

In this model, there is a smooth matching with the parton shower, since a soft gluon with $k_\perp$ smaller than the inverse string width will have no effect on the hadronic final state.

In summary, the string model has a very strong physical motivation. It is a model of confinement, unlike the earlier independent fragmentation models. It is universal and gives the best description of data. However, for many of the effects for which it gives a strongly motivated qualitative prediction, in practice its quantitative prediction depends on free parameters that can be tuned to data. The smooth matching to the parton shower can also be seen as a disadvantage if one wishes to learn about the perturbative phase of QCD evolution as it, in a sense, can cover up the precise information from the parton shower. This motivated people to think of a new model which will be discussed in the following sections.

3.4 Preconfinement and the Cluster Model

In the planar, or large N_c, approximation, a gluon is a colour-anticolour pair. One can follow the colour structure of the parton shower and find for each external parton its colour partner to which it is colour connected. One finds that these colour-singlet pairs tend to end up close in phase space. The mass spectrum of colour-singlet pairs is asymptotically independent of energy or the production process and is peaked at low mass $\sim Q_0$. It depends on Q_0 and Λ, but not the shower scale Q. This property is known as preconfinement and is the inspiration for the cluster hadronization model.

The cluster model is motivated by thinking about the spectrum of mesonic states constructed from given quark and antiquark flavours. The lightest states are narrow, but the heavier ones are broad resonances – above 1.5 GeV or so one can picture a continuum of overlapping states of different spins. One can then

think of the colour-anticolour pairs of preconfinement being projected directly onto this continuum. We call them clusters. These decay to lighter well-known resonances and stable hadrons. Once we have summed over all possible spins for a given process we effectively wash out all of the spin information and this assumption tells us that the decay should happen according to pure phase space. One immediate consequence of this is that heavier hadrons are suppressed – you get baryon and strangeness suppression "for free" (i.e. they are untuneable). The hadron-level properties are then fully determined by the cluster mass spectrum, i.e. by perturbative parameters. Q_0 is therefore a crucial parameter of the model.

This naïve cluster model works well for the bulk of colour singlet states but, although the cluster mass spectrum is peaked at small mass, there is a broad tail to high masses. A small fraction of clusters are too heavy for isotropic two-body decay to be a reasonable approximation. In the cluster fission model, these high mass colour-anticolour pairs split into two lighter clusters in a longitudinal, i.e. rather string-like, way. The fission threshold becomes another crucial parameter for tuning the cluster model as, although only $\sim$15 % of primary clusters get split, $\sim$50 % of hadrons come from them.

The cluster model was found to describe data reasonably well, with far fewer parameters than the string model. However, although it was found to work well for the majority of hadrons, it was noticed that the leading hadrons were not hard enough. This was cured, at the expense of an additional parameter, by saying that perturbatively-produced quarks remember their direction somewhat, with probability distribution

$$P(\theta^2) \sim \exp\left(-\theta^2/\theta_0^2\right), \tag{34}$$

so that this cluster fragments more along its own axis. This again is not completely isotropic but more string-like as it remembers the direction along which the colour is expanding. It also has more adjustable parameters to fit the data.

The founding philosophy of the string and cluster model are quite opposite. The *cluster model* emphasizes the perturbative phase of parton evolution and postulates that if this is correctly described, "any old model" of hadronization will be good enough. The *string model* emphasizes the non-perturbative dynamics of the confinement of partons and started initially from a very simple treatment of the production of those partons. The accumulation of more precise data have led the models to converge to something more similar. The string model has had successively refined perturbative evolution and the cluster model has become successively more string-like. This leads one to wonder whether nature is pointing us towards a model in which the flavour mix is largely determined by the perturbative dynamics, as in the cluster model, and their distributions largely determined by non-perturbative string dynamics.

We close this section by commenting briefly on the universality of hadronization parameters. With so many free parameters, one might question the predictive power of these models. However, one finds in practice that the parameters are universal: that a single set of parameters describes the data at a wide range of

energies and processes. One can show that this is a consequence of preconfinement: the perturbative production and evolution of partons takes care of the process- and energy-dependence and the transition from partons to hadrons is a local process, independent of these factors. Thus, hadronization models tuned to e^+e^- annihilation, and lower energy hadron collider data, are highly predictive for LHC events.

3.5 Secondary Decays

An often underestimated ingredient of event generators is the model of secondary particle decays. This is more important than often realized because in both the string and the cluster models it is rare that the clusters decay directly to the stable pions and kaons seen in the detector. Mostly they decay to higher resonances which then decay further. One might say that these decays have been measured and one can just "use the PDG", but often not all resonances in a given multiplet have been measured, and rarely do the measured branching fractions add up to 100 % or respect expected symmetries such as isospin exactly. So when these data tables in the MC are built, a lot of choices need to be made. Moreover, in the case of multi-body decays, the matrix elements are highly non-trivial and appropriate models have to be constructed for them. The decay tables, and the decay models that implement them, actually have a significant effect on the hadron yields, transverse momentum release and hadronization correction to event shapes. The choice of decay table should therefore be considered as part of the tuned parameter set and a change of decay tables should be accompanied by a re-tune.

4 Underlying Event

The preceding steps of hard process, parton shower, hadronization and secondary decays are sufficient to fully describe the final state of the hard process, in which a high energy parton from each incoming hadron interact to produce an arbitrarily complex final state. However, this process involves the extraction of a coloured parton from each of the hadrons, which are colourless bound states of many coloured partons. We therefore have to consider how the hadron remnants evolve, hadronize and, potentially, interact with each other.

In a proton's rest frame, it is a spherically symmetric extended object. Therefore in the laboratory frame where two protons collide at high energy, they look like extremely flattened discs due to Lorentz contraction, as shown in Fig. 15. Internal interactions are also extremely time dilated, so during the time that the discs overlap the protons' internal dynamics are effectively frozen. On the one hand, this means that a high energy interaction is extremely localized, and the whole of the parton shower and hadronization of the primary interaction happens in a very small

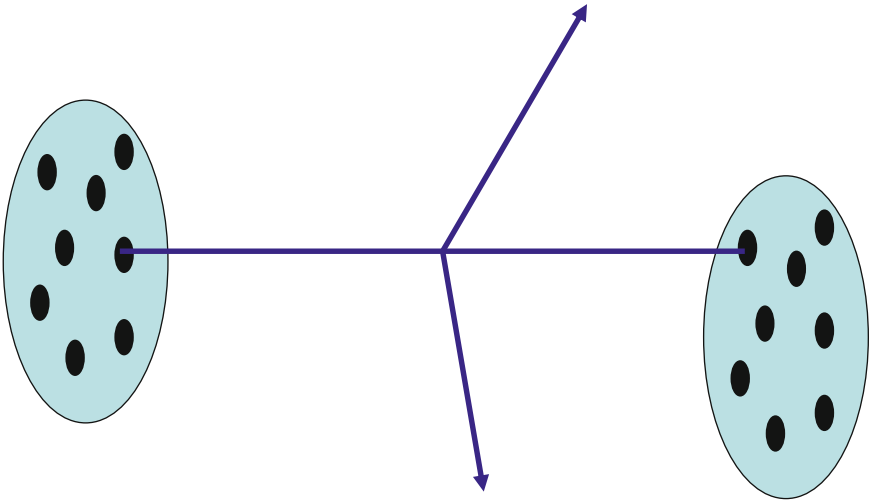

Fig. 15 Sketch of a proton-proton collision showing the Lorentz contraction of the protons

space-time region and there is not time for it to be affected by the rest of the proton. On the other hand, it means that there is a very large overlap between the other partons in the protons and the possibility of additional interactions.

Historically, there are two main models that have been used. Even though the first is effectively ruled out by Tevatron and LHC data, it is still useful to discuss, to draw out the contrasting features of the second, more successful, model.

The *non-perturbative model* is motivated by the fact that the soft parton-parton cross section is so large that there are many interactions everywhere in these discs and the assumption that these interactions are coherent across the discs. Thus the whole of one remnant interacts non-perturbatively with the whole of the other remnant. In the absence of an understanding of non-perturbative dynamics, our best hope is to parametrize data on these interactions. The only predictivity comes from the assumption that the underlying event at a given energy is independent of the hard process it underlies. This model was the default in HERWIG and is made available as an option in Herwig++ as it is still interesting to have a "straw man model" of soft hadronic interactions without any hard component. However, all the models that successfully describe the LHC data have a perturbative origin.

In the *perturbative models*, the idea is that the perturbative parton-parton cross section is so large that additional *local* parton-parton interactions between other partons in the proton dominate. We do not therefore have a coherent scattering but multiple independent parton-parton interactions distributed across the disc, each producing their own hard processes and parton showers as well as the initial one that we started with.

The underlying event is closely linked with what are often called "minimum bias" events. These are the final states of a typical proton-proton collision and typically consist of a small number of hadrons at low transverse momentum distributed

across a wide range of rapidities. Although the name "minimum bias" is widely used it is important to keep in mind that this is an experimental statement. By "minimum" we mean "as little as possible" so the amount of bias is dependent on the experiment. We would like to compare them with models that we think of as having zero bias, that are predicting all inelastic proton-proton collisions. To avoid confusion about the experiment-dependence of any minimum bias definition, the recent recommendation [3] is to describe the event class as "soft inclusive" events, reserving the name "minimum bias" for experimental attempts to measure these events.

In analysis, people often assume that they can remove the effect of the underlying event by measuring soft inclusive events and then subtracting these off, as the features of the two are very similar. This works reasonably well as a first approximation but if you look into the details, fluctuations in the amount of underlying event and correlations between the underlying event and the measured jets are extremely important. Making this assumption can potentially be quite dangerous and it is possible to underestimate the size of underlying event corrections.

Most jet cross sections are very steep, typically falling like the 5th or 6th power of p_t. If jets get a little extra energy from the underlying event their distribution gets shifted sideways, but since the distribution is so steeply falling a small shift sideways corresponds to a very large upwards or downwards shift in the curve. So a small contamination from the underlying event can give a large change in the jet production rate with given kinematics. This means that jet cross sections are sensitive to rare fluctuations in the underlying event and just subtracting off an average amount of underlying event is not necessarily meaningful. Processes with different p_t distributions will have different underlying event corrections – the steeper the p_t distribution is, the more a jet sample will be populated by lower p_t jets that have been shifted up by rare fluctuations in the underlying event. It is therefore extremely important to have reliable underlying event models that can predict this.[10] The way to avoid this trouble is to not tune to the average amount of underlying event but to correct on an event-by-event basis, this way fluctuations and correlations will be better taken into account [32].

4.1 Multiparton Interaction Model

The starting point for the perturbative model is the observation that the hard parton-parton cross section is so large that we have many parton-parton collisions in one proton-proton collision. This is demonstrated in Fig. 16, which shows three curves

[10]The underlying event itself is not usually assumed to be correlated with the process but there is a trigger bias – if you look at jets in a given kinematic range the distribution of the primary jets determines how much they are affected by the underlying event. For example, in Z and ZZ production not much difference in underlying event is expected, but because the ZZ process has a harder p_t distribution of accompanying jets, it is less affected by the underlying event.

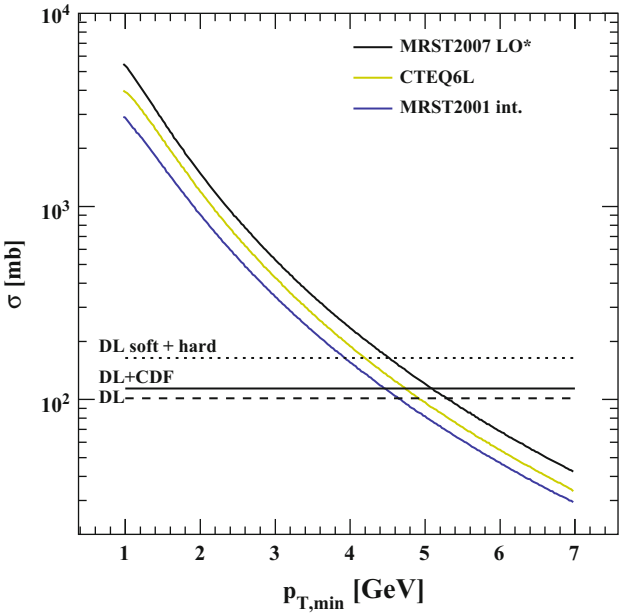

Fig. 16 Cross section at 14 TeV as a function of minimum p_t for different PDF sets (Reproduced from [33])

using different PDF sets, with their different α_s values. The total proton-proton cross section predicted by three different models is shown for comparison.[11]

For small $p_{t,min}$ and high energy the inclusive parton-parton cross section is larger than the total proton-proton cross section allowing more than one parton-parton scatter per proton-proton collision [34]. From PDFs calculated from deep inelastic scattering measurements, we know the distribution of momentum fractions of partons in the proton. What needs to be added is a model to describe the spatial distribution of partons within a proton. This is the only additional non-perturbative ingredient we need: with this we can calculate the distribution of number of scatters there are per proton-proton collision.

For these matter distributions, the current models usually make the assumption that x and b factorize:

$$n_i(x, b; \mu^2, s) = f_i(x; \mu^2)\, G(b, s), \tag{35}$$

[11] This figure was made for 14 TeV centre-of-mass energy and before the LHC measurements of the total cross section. The measured values at 7 and 8 TeV are closest to those predicted by the model labelled as "DL+CDF" in the figure.

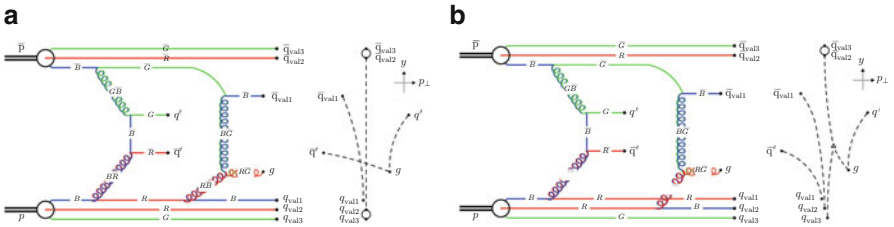

Fig. 17 Example of colour correlations for two $p\bar{p}$ events with the same hard scatter colour structure (Reproduced from [3], adapted from [35])

with $f_i(x;\mu^2)$ the usual (inclusive) parton distribution functions, and that n-parton distributions are independent:

$$n_{i,j}(x_i, x_j, b_i, b_j) = n_i(x_i, b_i)\, n_j(x_j, b_j), \tag{36}$$

etc. In these approximations, the number of scatters at a fixed impact parameter, i.e. with a given overlap between the two protons, is then given by a Poisson distribution. This can be integrated over impact parameter to calculate the n-scatter cross section:

$$\sigma_n = \int \mathrm{d}^2 b\, \frac{\left(A(b)\sigma^{inc}\right)^n}{n!}\, \exp\left(-A(b)\sigma^{inc}\right), \tag{37}$$

with

$$A(b) = \int \mathrm{d}^2 b_1\, G(b_1)\, \mathrm{d}^2 b_2\, G(b_2)\, \delta^{(2)}(\mathbf{b} - \mathbf{b}_1 + \mathbf{b}_2). \tag{38}$$

These ingredients are sufficient to generate a number of scatters and their kinematics and parton showers. The one remaining complication is how the colour of the different scatters is connected to each other. This has been studied by the Pythia authors in some detail [35] and recently also by the Herwig authors [36, 37]. Figure 17 shows two $p\bar{p}$ events for which the hard processes ($gg \to q\bar{q}$ and $qg \to qg$) have exactly the same colour structure but the colour connections between the scatters and the external protons is different in the two cases. Although the parton showering will be identical in the two cases, the hadronization will differ, because the string connections (represented by the dashed lines in the right-hand part of each figure) differ.

Although perturbation theory can specify everything about the colour connections in the centre of the event, it doesn't tell us how these colours are hooked into the wavefunction of the original protons. So we need to supplement our model. The Pythia authors have studied different algorithms to do this in some detail and identified experimental observables that help to constrain them [38].

4.2 The Herwig++ Multiparton Interaction and Colour Reconnection Models

The multiparton interaction model in Herwig++ is developed from the one available as a plug-in to HERWIG, called Jimmy [39], but with a number of new features [36, 40]. The idea is to use the eikonal model and optical theorem to connect the partonic scattering cross sections to the total, inelastic and elastic hadronic cross sections. A simple separation of partonic scatters is made into hard (above $p_{t,min} \sim$ 3–5 GeV), distributed perturbatively around 'hot spots' of high parton density in the protons, and soft (below $p_{t,min}$), with a simple distribution with Gaussian[12] transverse momentum and valence-like momentum fraction distributed across the whole of the protons' radii. Once the total cross section and elastic form factor are fixed by data there are only two free parameters: $p_{t,min}$ and the effective hot spot radius.

When first implemented, this model gave a good description of underlying event data, but failed badly for soft inclusive analyses. It was realized that this was due to the issue of colour correlations between the scatters, which was not very carefully handled in the first implementation. Röhr, Siodmok and Gieseke have implemented a new model of colour reconnections in Herwig++ [37] based on the momentum structure. This also gives reconnection effects in e^+e^- annihilation, so a refit of LEP-I and LEP-II data is necessitated, but the conclusion is that one can get a good fit of the e^+e^- and LHC data, see for example Fig. 18 where the red line is the best fit to the data without colour reconnections and is clearly nothing like the data. An important conclusion of this study is that the hadronization parameters are correlated with the reconnection probability – changing one necessitates a retuning of the other, but a good fit can be obtained for a wide range of colour reconnection probabilities.

4.3 Pythia Implementation

Pythia was the first event generator to incorporate multiparton interactions and its implementation is very well developed. An interesting feature that has emerged, that is not shared by Herwig or Sherpa, is the possibility that through colour connection effects there can be an interplay between multiparton interactions and the initial-state parton shower [41]. Starting with a single hard interaction at some value of p_t, the simulation evolves downwards in p_t, with the possibility at every step of either

[12] In fact, once the parameters are fixed, an interesting feature emerges: the width-squared of the Gaussian is forced to be negative, giving an "inverted Gaussian" with very few events at very low transverse momentum and a concentration of events around $p_{t,min}$. This lends support to the multiparton interaction model and the idea that the entire cross section could be described perturbatively at high energy.

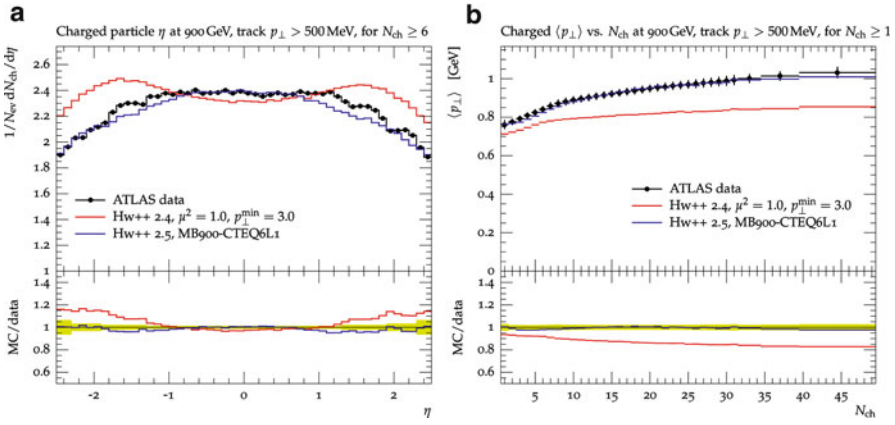

Fig. 18 Comparison of Herwig++ 2.4.2, without colour reconnections, and Herwig++ 2.5, with colour reconnections, to ATLAS minimum-bias distributions at $\sqrt{s} = 0.9\,\text{TeV}$ with $N_{ch} \geq 6$, $p_t > 500\,\text{MeV}$ and $|\eta| < 2.5$ (Reproduced from [37])

generating an emission from an incoming parton or an additional scatter. At any point during this evolution there can be colour cross-talk between these different interactions and this will affect the distribution of hadronization and of the partons that are radiated by these multiple scatters.

A recent study [42] also considered rescattering, where two partons out of one proton can scatter with the same parton from the other. This is suppressed by the fact that it must be a local process: the two partons in one proton must both be overlapping with the other. Nevertheless it does give another contribution to the fluctuations in the underlying event.

Most existing models make the simplifying assumption in Eq. (35) of factorization of x and b, i.e. of the momentum and spatial distribution of partons. In another recent study [43], Corke and Sjöstrand implemented a model without this assumption by considering a Gaussian matter distribution with an x-dependent width,

$$a(x) = a_0 \left(1 + a_1 \ln \frac{1}{x}\right), \tag{39}$$

and looked at what you could learn about these parameters from the data. The effect is to start producing more correlation with the underlying event. A higher mass final state is produced at higher x, therefore it has a narrower matter distribution and more underlying event. There is a correlation between the momentum used in the hard collision and the underlying event that accompanies it. They compared the underlying event in Z events and in events that produce a Z' of 1 TeV and found significant differences. They could obtain equally good fits of the existing underlying event data but with significant differences in their extrapolation to

higher-scale processes. This is clearly something that requires further study, to improve the models and to understand the uncertainty they introduce in high-mass searches using jets, for example.

4.4 Underlying Event Measurements

Despite a ∼25 year history, many aspects of our understanding of multi-parton interactions are still in their infancy. The Tevatron and especially LHC experiments have already opened up huge areas for further study, not only with a big step up in energy but also with much higher efficiency, purity and phase space coverage than the previous measurements. There has also been a big change in the culture around the measurements, with an emphasis on physical (experiment-independent and generator-independent) observables that can be directly compared between experiments and with a wide variety of models, now or in the future. There is also a move towards making more targeted measurements of observables that are sensitive to specific physical effects, such as colour reconnections. The general conclusion is that all the existing models can describe the general underlying event and soft inclusive data well with tuning. The emphasis is moving towards understanding of correlations between hard and underlying events, rare corners of phase space (such as high multiplicity soft events) and the relationships between different model components. One of the main motivations for these studies is the fact that jet corrections depend strongly on these correlations, and high moments of distributions, and are physics-process dependent. A deeper understanding, and greater predictivity, is still needed.

5 Summary

As a summary of our discussion of event generators, we recall the main subjects that we have covered, commenting on how well they are understood from first principles. We briefly touched on the hard process which is generally a direct implementation of tree-level perturbation theory and hence extremely well understood. We discussed in detail the parton shower which is an approximation to all-order perturbation theory and therefore in principle well understood. Various approximations are made in constructing parton showers and the effect of these is not always as small as anticipated. The cutting edge here is the matching between higher order fixed-order perturbation theory and parton showers, which should, in principle, be fully understandable from perturbation theory, but is at present the subject of some uncertainty. We then talked about hadronization which is less well understood from first principles. Although there are different models, they are well constrained by data and the extrapolation to LHC data is considered to be fairly reliable. Lastly, the underlying event is the least well understood out of all these. It is only

weakly constrained by previous data and different models that fit the available data give quite different extrapolations. Moreover, it is important to recall that correlations and rare fluctuations in the underlying event are as important as its average properties and are even less well tied down.

Monte Carlo event generators are increasingly used as tools in almost every aspect of high energy collider physics. As the data become more precise it becomes increasingly important not to use them as black boxes, but to question how reliable they are for the application at hand. The important question to ask is "What physics is dominating my effect?". We hope that these lecture notes have helped equip the reader to answer this question.

Acknowledgements We thank the organizers of the 69th Scottish Universities' Summer School in Physics for the invitation to give these lectures, for organizing a very enjoyable and stimulating school, and for the putting competition, whisky tasting and ceilidh. The lectures have been developed from similar courses given over the years at several summer schools and MHS thanks members of the CTEQ and MCnet collaborations in particular for their feedback and encouragement.

References

1. G. Aad et al., ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012) [arXiv:1207.7214 [hep-ex]]
2. S. Chatrchyan et al., CMS Collaboration, Phys. Lett. B **716**, 30 (2012)
3. A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss, L. Lönnblad et al., Phys. Rep. **504**, 145 (2011)
4. R.K. Ellis, W.J. Stirling, B.R. Webber, QCD and collider physics. Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996)
5. F. Abe et al., CDF Collaboration, Phys. Rev. D **50**, 5562 (1994)
6. S. Catani, S. Dittmaier, M.H. Seymour, Z. Trocsanyi, Nucl. Phys. B **627**, 189 (2002)
7. E. Norrbin, T. Sjöstrand, Nucl. Phys. B **603**, 297 (2001)
8. Z. Nagy, D.E. Soper, JHEP **0510**, 024 (2005)
9. M. Dinsdale, M. Ternick, S. Weinzierl, Phys. Rev. D **76**, 094003 (2007)
10. S. Schumann, F. Krauss, JHEP **0803**, 038 (2008)
11. J.-C. Winter, F. Krauss, JHEP **0807**, 040 (2008)
12. W.T. Giele, D.A. Kosower, P.Z. Skands, Phys. Rev. D **78**, 014026 (2008)
13. S. Catani, M.H. Seymour, Phys. Lett. B **378**, 287 (1996); S. Catani, M.H. Seymour, Nucl. Phys. B **485**, 291 (1997) [Erratum-ibid. B **510**, 503 (1998)]
14. D.A. Kosower, Phys. Rev. D **57**, 5410 (1998)
15. S. Catani, F. Krauss, R. Kuhn, B.R. Webber, JHEP **0111**, 063 (2001)
16. F. Krauss, JHEP **0208**, 015 (2002)
17. L. Lönnblad, JHEP **0205**, 046 (2002)
18. J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M.L. Mangano, M. Moretti et al., Eur. Phys. J. C **53**, 473 (2008)
19. S. Frixione, B.R. Webber, JHEP **0206**, 029 (2002)
20. P. Nason, G. Ridolfi, JHEP **0608**, 077 (2006)
21. N. Lavesson, L. Lönnblad, JHEP **0812**, 070 (2008)
22. K. Hamilton, P. Nason, JHEP **1006**, 039 (2010)
23. S. Höche, F. Krauss, M. Schönherr, F. Siegert, JHEP **1108**, 123 (2011)

24. T. Sjöstrand, P. Eden, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna, E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001)
25. G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, B.R. Webber, JHEP **0101**, 010 (2001)
26. L. Lönnblad, Comput. Phys. Commun. **71**, 15 (1992)
27. T. Sjöstrand, S. Mrenna, P.Z. Skands, JHEP **0605**, 026 (2006)
28. T. Sjöstrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008)
29. T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, J. Winter, JHEP **0902**, 007 (2009)
30. M. Bähr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson et al., Eur. Phys. J. C **58**, 639 (2008)
31. S. Plätzer, S. Gieseke, Eur. Phys. J. C **72**, 2187 (2012)
32. M. Cacciari, G.P. Salam, Phys. Lett. B **659**, 119 (2008)
33. M. Bähr, J.M. Butterworth, M.H. Seymour, JHEP **0901**, 065 (2009)
34. T. Sjöstrand, M. van Zijl, Phys. Rev. D **36**, 2019 (1987)
35. T. Sjöstrand, P.Z. Skands, JHEP **0403**, 053 (2004)
36. M. Bähr, S. Gieseke, M.H. Seymour, JHEP **0807**, 076 (2008)
37. S. Gieseke, C.A. Röhr, A. Siódmok, Eur. Phys. J. C **72**, 2225 (2012)
38. P.Z. Skands, D. Wicke, Eur. Phys. J. C **52**, 133 (2007)
39. J.M. Butterworth, J.R. Forshaw, M.H. Seymour, Z. Phys. C **72**, 637 (1996)
40. I. Borozan, M.H. Seymour, JHEP **0209**, 015 (2002)
41. T. Sjöstrand, P.Z. Skands, Eur. Phys. J. C **39**, 129 (2005)
42. R. Corke, T. Sjöstrand, JHEP **1001**, 035 (2010)
43. R. Corke, T. Sjöstrand, JHEP **1105**, 009 (2011)

Statistics for Searches at the LHC

Glen Cowan

Abstract This chapter describes several topics in statistical data analysis as used in High Energy Physics. It focuses on areas most relevant to analyses at the LHC that search for new physical phenomena, including statistical tests for discovery and exclusion limits. Particular attention is given to the treatment of systematic uncertainties through nuisance parameters.

1 Introduction

The primary goal of data analysis in High Energy Physics (HEP) is to test our understanding of particle interactions and, in doing so, to search for phenomena that go beyond the existing framework of the Standard Model. These lectures describe some of the basic statistical tools needed to do this.

Despite efforts to make the lectures self contained, some familiarity with basic ideas of statistical data analysis is assumed. Introductions to the subject can be found, for example, in the reviews of the Particle Data Group [1] or in the texts [2–6].

Brief reviews are given of probability in Sect. 2 and frequentist hypothesis tests in Sects. 3 and 4. These are applied to establishing discovery and setting limits (Sect. 5) and are extended using the profile likelihood ratio (Sect. 6), from which one can construct unified intervals (Sect. 8). Bayesian limits are discussed in Sect. 9 and all of the methods for limits are illustrated using the example of a Poisson counting experiment in Sect. 10. Application of the standard tools for discovery and limits leads to a number of difficulties, such as exclusion of models to which one has no sensitivity (Sect. 11) and the look-elsewhere effect (Sect. 12). Section 13 illustrates how the methods have been applied in the search for the Higgs boson. In Sect. 14 we examine why one traditionally requires five-sigma significance to claim a discovery and finally some conclusions are drawn in Sect. 15. The lectures

G. Cowan (✉)
Physics Department, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK
e-mail: g.cowan@rhul.ac.uk

E. Gardi et al. (eds.), *LHC Phenomenology*, Scottish Graduate Series,
DOI 10.1007/978-3-319-05362-2_9

as presented at SUSSP also included material on unfolding or deconvolution of measured distributions which is not included here but can be found in Ref. [7] and Chapter 11 of Ref. [2].

2 Review of Probability

When analyzing data in particle physics one invariably encounters uncertainty, at the very least coming from the intrinsically random nature of quantum mechanics. These uncertainties can be quantified using probability, which was defined by Kolmogorov [8] using the language of set theory. Suppose a set S contains elements that can form subsets $A, B, \ldots$. As an example, the elements may represent possible outcomes of a measurement but here we are being abstract and we do not need to insist at this stage on a particular meaning. The three axioms of Kolmogorov can be stated as

1. For all $A \subset S$, there is a real-valued function P, the probability, with $P(A) \geq 0$;
2. $P(S) = 1$;
3. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

In addition we define the conditional probability of A given B (for $P(B) \neq 0$) as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

From these statements we can derive the familiar properties of probability. They do not, however, provide a complete recipe for assigning numerical values to probabilities nor do they tell us what these values mean.

Of the possible ways to interpret a probability, the one most commonly found in the physical sciences, is as a limiting frequency. That is, we interpret the elements of the sample space as possible outcomes of a measurement, and we take $P(A)$ to mean the fraction of times that the outcome is in the subset A in the limit where we repeat the measurement an infinite number of times under "identical" conditions:

$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is in } A}{n}. \tag{2}$$

Use of probability in this way leads to what is called the *frequentist* approach to statistics. Probabilities are only associated with outcomes of repeatable observations, not to hypothetical statements such as "supersymmetry is true". Such a statement is either true or false, and this will not change upon repetition of any experiment.

Whether SUSY is true or false is nevertheless uncertain and we can quantify this using probability as well. To define what is called *subjective probability* one

interprets the elements of the set S as *hypotheses*, i.e., statements that are either true or false, and one defines,

$$P(A) = \text{degree of belief that } A \text{ is true.} \tag{3}$$

Use of subjective probability leads to what is called *Bayesian statistics*, owing to its important use of Bayes' theorem described below.

Regardless of its interpretation, any quantity that satisfies the axioms of probability must obey Bayes' theorem, which states,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{4}$$

This can be derived from the definition of conditional probability (1), which we can write as $P(A \cap B) = P(B)P(A|B)$, or equivalently by changing labels as $P(B \cap A) = P(A)P(B|A)$. These two probabilities are equal, however, because $A \cap B$ and $B \cap A$ refer to the same subset. Equating them leads directly to Eq. (4).

In Bayes' theorem (4) the condition B represents a restriction imposed on the sample space S such that anything outside of B is not considered. If the sample space S can be expressed as the union of some disjoint subsets A_i, $i = 1, 2, \ldots$, then the factor $P(B)$ appearing in the denominator can be written $P(B) = \sum_i P(B|A_i)P(A_i)$ so that Bayes' theorem takes on the form,

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}. \tag{5}$$

In *Bayesian* (as opposed to frequentist) statistics, one uses subjective probability to describe one's degree of belief in a given theory or hypothesis. The denominator in Eq. (5) can be regarded as a constant of proportionality and therefore Bayes' theorem can be written as,

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}), \tag{6}$$

where "theory" represents some hypothesis and "data" is the outcome of the experiment. Here $P(\text{theory})$ is the *prior* probability for the theory, which reflects the experimenter's degree of belief before carrying out the measurement, and $P(\text{data}|\text{theory})$ is the probability to have gotten the data actually obtained, given the theory, which is also called the *likelihood*.

Bayesian statistics provides no fundamental rule for obtaining the prior probability; in general this is subjective and may depend on previous measurements, theoretical prejudices, etc. Once this has been specified, however, Eq. (6) tells how the probability for the theory must be modified in the light of the new data to give the *posterior* probability, $P(\text{theory}|\text{data})$. As Eq. (6) is stated as a proportionality, the probability must be normalized by summing (or integrating) over all possible hypotheses.

3 Hypothesis Tests

One of the fundamental tasks in a statistical analysis is to test whether the predictions of a given model are in agreement with the observed data. Here we will use $\mathbf{x}$ to denote the outcome of a measurement; it could represent a single quantity or a collection of values. A hypothesis H means a statement for the probability to find the data $\mathbf{x}$ (or if $\mathbf{x}$ includes continuous variables, H specifies a probability density function or pdf). We will write $P(\mathbf{x}|H)$ for the probability to find data $\mathbf{x}$ under assumption of the hypothesis H.

Consider a hypothesis H_0 that we want to test (we will often call this the "null" hypothesis) and an alternative hypothesis H_1. In frequentist statistics one defines a *test* of H_0 by specifying a subset of the data space called the *critical region*, w, such that the probability to observe the data there satisfies,

$$P(\mathbf{x} \in w|H_0) \leq \alpha. \tag{7}$$

Here α is a constant specified before carrying out the test, usually set by convention to a small value such as 5 %. For continuous data, one takes the relation above as an equality. If the data are discrete, such as a number of events, then there may not exist any subset of the data values whose summed probability is exactly equal to α, so one takes the critical region to have a probability up to α. The critical region w defines the test. If the data are observed in w then the hypothesis H_0 is rejected.

Up to this point the sole defining property of the test is Eq. (7), which states that the probability to find the data in the critical region is not more than α. But there are in general many if not an infinite number of possible subsets of the data space that satisfy this criterion, and it is not clear which should be taken as the critical region. This is where the alternative hypothesis H_1 comes into play. One would like the critical region to be chosen such that there is as high a probability as possible to find the data there if the alternative is true, while having only the fixed probability α assuming H_0, as illustrated schematically in Fig. 1.

Rejecting the hypothesis H_0 if it is in fact true is called a type I error. By construction, the probability for this to occur is the size of the test, α. If on the other hand we do not reject H_0 but we should have, because the alternative H_1 was true, then this is called a type II error. The probability to reject the null hypothesis if the alternative H_1 is true is called the *power* of the test with respect to H_1, which is one minus the probability of a type II error.

A *significance test* of a hypothesis H is closely related to the tests described above. Suppose a measurement results in data $\mathbf{x}$ (a single number or a collection of many values) for which the hypothesis H predicts the probability $P(\mathbf{x}|H)$. We observe a single instance of $\mathbf{x}$, say, $\mathbf{x}_{\text{obs}}$, and we want to quantify the level of agreement between this outcome and the predictions of H.

To do this the analyst must specify what possible data values would constitute a level of incompatibility with H that is equal to are greater than that between H and the observed data $\mathbf{x}_{\text{obs}}$. Once this is given, then one computes the p-value of H as the probability, under assumption of H, to find data in this region of equal or greater incompatibility.

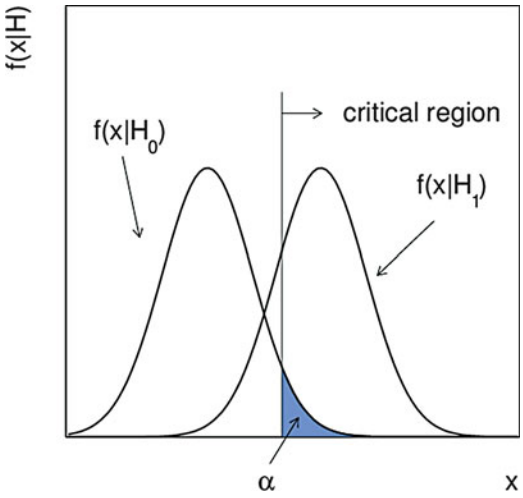

Fig. 1 Illustration of the critical region of a statistical test (see text)

When computing the p-value there is clearly some ambiguity as to what data values constitute greater incompatibility with H than others. When we say that a given $\mathbf{x}$ has less compatibility with H, we imply that it has more compatibility with some alternative hypothesis. This is analogous to the ambiguity we encountered in determining the critical region of a test.

We can see this connection more directly by using the p-value to specify the critical region for a test of H_0 of size α as the set of data values that would have a p-value less than or equal to α. The resulting test will have a certain power with respect to any given alternative H_1, although these may not have been used explicitly when constructing the p-value.

In a frequentist test, we *reject* H_0 if the data are found in the critical region, or equivalently, if the p-value of H_0 is found less or equal to α. Despite this language, it is not necessarily true that we would then believe H_0 to be false. To make this assertion we should quantify our degree of belief about H_0 using subjective probability as described above, and it must be computed using Bayes' theorem:

$$P(H_0|\mathbf{x}) = \frac{P(\mathbf{x}|H_0)\pi(H_0)}{\sum_i P(\mathbf{x}|H_i)\pi(H_i)}. \tag{8}$$

As always, the posterior $P(H_0|\mathbf{x})$ is proportional to the prior $\pi(H_0)$, and this would need to be specified if we want to express our degree of belief that the hypothesis is true.

It is also important to note that the p-value of a hypothesis H_0 is not the same as the probability (8) that it is true, but rather the probability, under assumption of H_0, to find data with at least as much incompatibility with H_0 as the data actually found. The p-value thus does not depend on prior probabilities.

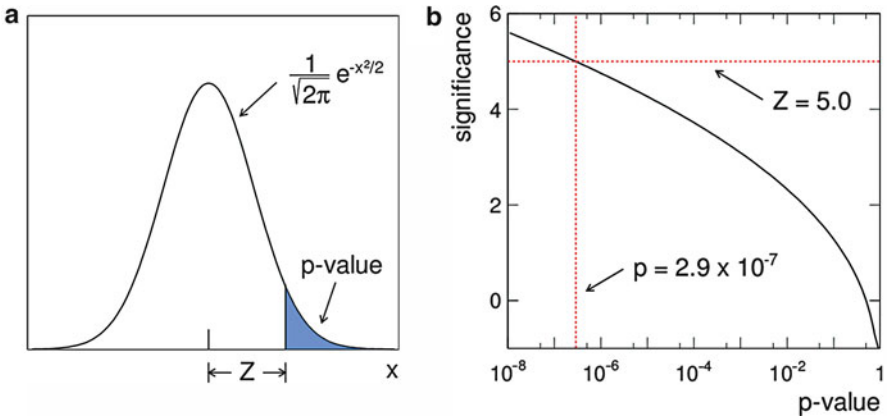

Fig. 2 (**a**) Illustration of the definition of significance Z and (**b**) the significance as function of the p-value

For most of these lectures we will stay within the frequentist framework. The result of our analysis will be a p-value for the different models considered. If this is less than some specified value α, we reject the model.

Often the p-value is translated into an equivalent quantity called the *significance*, Z, defined by,

$$Z = \Phi^{-1}(1 - p). \tag{9}$$

Here Φ is the cumulative standard Gaussian distribution (zero mean, unit variance) and Φ^{-1} is its inverse function, also called the *quantile* of the standard Gaussian. The definition of significance is illustrated in Fig. 2a and the significance versus p-value is shown in Fig. 2b. Often a significance of $Z = 5$ is used as the threshold for claiming discovery of a new signal process. This corresponds to a very low p-value of 2.9×10^{-7} for the no-signal hypothesis. The rationale for using such a low threshold is discussed further in Sect. 14.

Although we can simply take Eq. (9) as our defining relation for Z, it is useful to compare to the case of measuring a quantity x that follows a Gaussian distribution with unknown mean μ. Suppose we want to test the hypothesis $\mu = 0$ against the alternative $\mu > 0$. In this case we would take the critical region of the test to contain values of x greater than a certain threshold, or equivalently, we would define the p-value to be the probability to find x as large as we found or larger. In this case the significance Z is simply the value of x observed, measured in units of its standard deviation σ. For this reason one often refers to finding a significance Z of, say, 2.0 as a *two*-sigma effect.

4 Choice of Critical Region, Test Statistics

We now examine more closely the question of how best to define the critical region of a test and for this suppose we want to select a sample of events of a desired type (signal, denoted s) and reject others that we do not want (background, b). That is, for each event we will measure some set of quantities $\mathbf{x}$, which could represent different kinematic variables such as the missing energy, number of jets, number of muons, and so forth. Then for each event carry out a test of the background hypothesis, and if this is rejected it means we select the event as a candidate signal event.

Suppose that the s and b hypotheses imply probabilities for the data of $P(\mathbf{x}|s)$ and $P(\mathbf{x}|b)$, respectively. Figure 3 shows these densities for two components of the data space along with possible boundaries for the critical region.

Figure 3a represents what is commonly called the 'cut-based' approach. One selects signal events by requiring $x_1 < c_1$ and $x_2 < c_2$ for some suitably chosen cut values c_1 and c_2. If x_1 and x_2 represent quantities for which one has some intuitive understanding, then this can help guide one's choice of the cut values.

Another possible decision boundary is made with a diagonal cut as shown in Fig. 3b. One can show that for certain problems a linear boundary has optimal properties, but in the example here, because of the curved nature of the distributions, neither the cut-based nor the linear solution is as good as the nonlinear boundary shown in Fig. 3c.

The decision boundary is a surface in the n-dimensional space of input variables, which can be represented by an equation of the form $y(\mathbf{x}) = y_{\text{cut}}$, where y_{cut} is some constant. We accept events as corresponding to the signal hypothesis if they are on one side of the boundary, e.g., $y(\mathbf{x}) \leq y_{\text{cut}}$ could represent the acceptance region and $y(\mathbf{x}) > y_{\text{cut}}$ could be the rejection region.

Equivalently we can use the function $y(\mathbf{x})$ as a scalar *test statistic*. Once its functional form is specified, we can determine the pdfs of $y(\mathbf{x})$ under both the signal and background hypotheses, $p(y|s)$ and $p(y|b)$. The decision boundary is now effectively a single cut on the scalar variable y, as illustrated in Fig. 4.

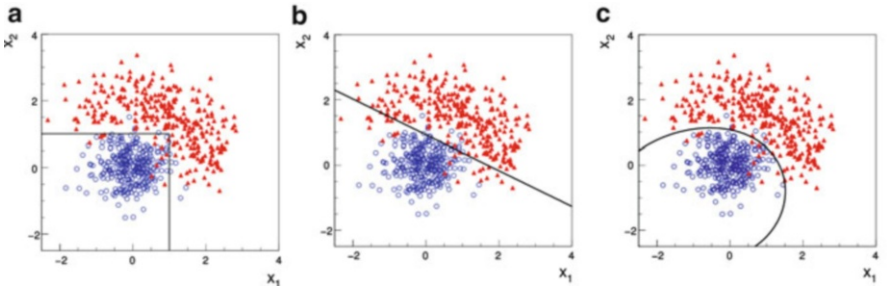

Fig. 3 Scatter plots of two variables corresponding to two hypotheses: background (H_0) and signal (H_1). The critical region for a test of H_0 could be based, e.g., on (**a**) cuts, (**b**) a linear boundary, (**c**) a nonlinear boundary

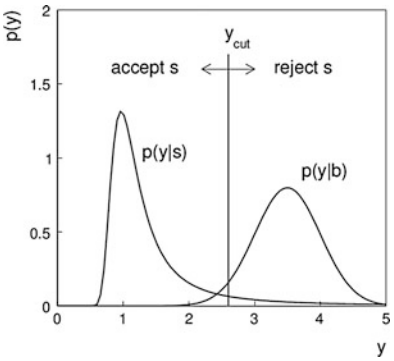

Fig. 4 Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses

We would like to design a test to have a high probability to reject a hypothesis if it is false, which is what we have called the power of the test. Unfortunately a test with maximum power with respect to one alternative will not be optimal with respect to others, so there is no such thing as an ideal "model-independent" test. Nevertheless, for a specific pair of signal and background hypotheses, it turns out that there is a well defined optimal solution to our problem. The *Neyman–Pearson* lemma (see, e.g., Ref. [9]) states that for a test of a given significance level of the background hypothesis (i.e., fixed background efficiency), one obtains the maximum power relative to the signal hypothesis (signal efficiency) by defining the critical region w such that for $\mathbf{x} \in w$ the *likelihood ratio*, i.e., the ratio of pdfs for signal and background,

$$y(\mathbf{x}) = \frac{f(\mathbf{x}|s)}{f(\mathbf{x}|b)}, \tag{10}$$

is greater than or equal to a given constant, and it is less than this constant everywhere outside the critical region. This is equivalent to the statement that the ratio (10) represents the test statistic with which one obtains the highest signal efficiency for a given background efficiency.

In principle the signal and background theories should allow us to work out the required functions $f(\mathbf{x}|s)$ and $f(\mathbf{x}|b)$, but in practice the calculations are too difficult and we do not have explicit formulae for these. What we have instead of $f(\mathbf{x}|s)$ and $f(\mathbf{x}|b)$ are complicated Monte Carlo programs, from which we can sample $\mathbf{x}$ to produce simulated signal and background events. Because of the multivariate nature of the data, where $\mathbf{x}$ may contain at least several or perhaps even hundreds of components, it is a nontrivial problem to construct a test with a power approaching that of the likelihood ratio.

In the usual case where the likelihood ratio (10) cannot be used explicitly, there exist a variety of other multivariate classifiers such as neural networks, boosted decision trees and support vector machines that effectively separate different

types of events. Descriptions of these methods can be found, for example, in the textbooks [10–13], lecture notes [14] and proceedings of the PHYSTAT conference series [15]. Software for HEP includes the `StatPatternRecognition` [16] and `TMVA` [17] packages.

5 Frequentist Treatment of Discovery and Limits

The use of a statistical test in a particle physics analysis involving different event types comes up in different ways. Sometimes both event classes are known to exist, and the goal is to select one class (signal) for further study. For example, top-quark production in proton–proton collisions is a well-established process. By selecting these events one can carry out precise measurements of the top quark's properties such as its mass. This was the basic picture in the previous section. The measured quantities referred to individual events, and we tested the hypothesized event type for each.

In other cases, the signal process could represent an extension to the Standard Model, say, supersymmetry, whose existence is not yet established, and the goal of the analysis is to see if one can do this. Here we will imagine the "data" as representing not individual events but a sample of events, i.e., an entire "experiment". If the signal process we are searching for does not exist, then our sample will consist entirely of background events, e.g., those due to Standard Model processes. If the signal does exist, then we will find both signal and background events. Thus the hypothesis we want to test is,

$$H_0 \ : \ \text{only background processes exist,}$$

versus the alternative,

$$H_1 \ : \ \text{both signal and background exist.}$$

We will refer to the hypothesis H_0 as the background-only model (or simply "b") and the alternative H_1 as the signal-plus-background model, $s + b$. The Neyman-Pearson lemma still applies. In a test of H_0 of a given size, the highest power relative to H_1 is obtained when the critical region contains the highest values of the likelihood ratio $L(H_1)/L(H_0)$. Here, however, the likelihood is the probability for the entire set of data from the experiment, not just for individual events.

Rejecting H_0 means in effect discovering a new phenomenon. Of course before we believe that we have made a new discovery, a number of other concerns must be addressed, such as our confidence in the reliability of the statistical models used, the plausibility of the new phenomenon and the degree to which it can describe the data. Here, however, we will simply focus on question of statistical significance and in effect equate "rejecting the background-only hypothesis" with "discovery".

Often in HEP one claims discovery when the p-value of the background-only hypothesis is found below 2.9×10^{-7}, corresponding to a 5-sigma effect. We will revisit the rationale behind this threshold in Sect. 14.

Even if one fails to discover new physics by rejecting the background-only model, one can nevertheless test various signal models and see whether they are compatible with the data. Signal models are usually characterized by some continuous parameters representing, e.g., the masses of new particles. If we carry out a test of size α for all possible values of the parameters, then those that are not rejected constitute what is called a *confidence region* for the parameters with a *confidence level* of $\mathrm{CL} = 1 - \alpha$. By construction a hypothesized point in parameter space will, if it is true, be rejected with probability α. Therefore the confidence region will contain the true value of the parameters with probability $1 - \alpha$. For purposes of confidence limits one typically uses a test of size $\alpha = 0.05$, which is to say the regions have a confidence level of 95 %.

If the problem has only one parameter, then the region is called a confidence interval. An important example is where a parameter μ is proportional to the cross section for the signal process whose existence is not yet established. Here one often wants to test a hypothetical value relative to the alternative hypothesis that the signal does not exist, i.e., $\mu = 0$. The critical region of the test is then taken have higher probability for the lower values of the parameter.

For example, suppose the data consist of a value x that follows a Gaussian distribution with unknown mean μ and known standard deviation σ. If we test a value μ relative to the alternative of a smaller value, then the critical region will consist of values of $x < c$ for some constant c such that,

$$\alpha = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx = \Phi\left(\frac{c-\mu}{\sigma}\right), \tag{11}$$

or,

$$c = \mu - \sigma \Phi^{-1}(1-\alpha). \tag{12}$$

If we take, e.g., $\alpha = 0.05$, then the factor $\Phi^{-1}(1-\alpha) = 1.64$ says that the critical region starts at 1.64 standard deviations below the value of μ being tested. If x is observed any lower than this, then the corresponding μ is rejected.

Equivalently we can take the p-value of a hypothesized μ, p_μ, as the probability to observe x as low as we found or lower, and we then reject μ if we find $p_\mu \leq \alpha$. The highest value of μ that we do not reject is called the *upper limit* of μ at a confidence level of $1 - \alpha$, and we will write this here as μ_{up}. Lower limits, μ_{lo}, can of course be constructed using an analogous procedure. In practice these points are found by setting $p_\mu = \alpha$ and solving for μ. There are a number of subtle issues connected with limits derived in this way and we will return to these in Sect. 11.

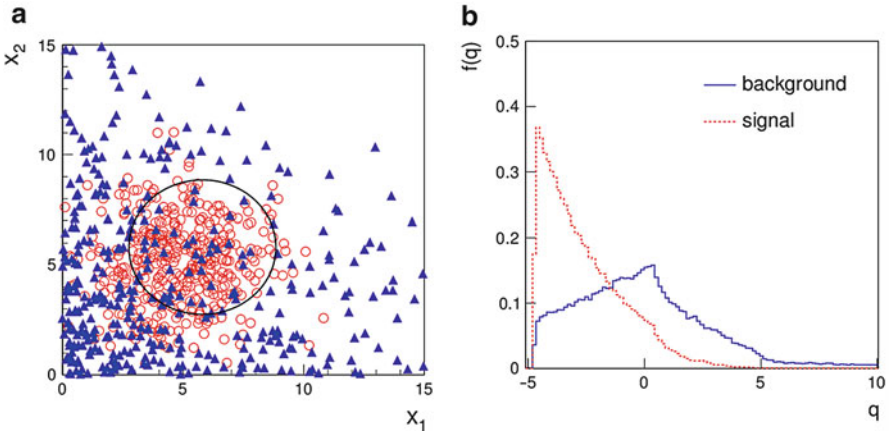

Fig. 5 (**a**) Distributions of $\mathbf{x} = (x_1, x_2)$ for events of type signal (*red circles*) and background (*blue triangles*) shown with a contour of constant likelihood ratio; (**b**) the distribution of the statistic q for signal and background events

5.1 A Toy Example

Consider an experiment where, for each selected event, we measure two quantities, which we can write as a vector $\mathbf{x} = (x_1, x_2)$. Suppose that for background events $\mathbf{x}$ follows,

$$f(\mathbf{x}|b) = \frac{1}{\xi_1} e^{-x/\xi_1} \frac{1}{\xi_2} e^{-x/\xi_2}, \tag{13}$$

and for a certain signal model they follow,

$$f(\mathbf{x}|s) = C \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1-\mu_1)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x_2-\mu_2)^2/2\sigma_2^2}, \tag{14}$$

where $x_1 \geq 0$, $x_2 \geq 0$ and C is a normalization constant. The distribution of events generated according to these hypotheses are shown in Fig. 5a.

First, suppose that the signal and background both correspond to event types that are known to exist and the goal is simply to select signal. In this case we can exploit the Neyman-Pearson lemma and base the selection on the likelihood ratio,

$$y(\mathbf{x}) = \frac{f(\mathbf{x}|s)}{f(\mathbf{x}|b)}. \tag{15}$$

We can define the same critical region by using any monotonic function of the likelihood ratio, and in this case it is useful to take

$$q = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - \frac{2x_1}{\xi_1} - \frac{2x_2}{\xi_2} = -2 \ln y(\mathbf{x}) + \text{const.} \tag{16}$$

Distributions of the statistic q for the background and signal hypotheses (13) and (14) are shown in Fig. 5b. This shows that a sample enhanced in signal events can be selected by selecting events with q less than a given threshold, say, q_{cut}.

Now suppose instead that the signal process is not known to exist and the goal of the analysis is to search for it. Suppose that the expected numbers events are b of background and s for a given signal model. For now assume that the model's prediction for both of these quantities can be determined with negligible uncertainty. The actual number of events n that we find can be modeled as a Poisson distributed quantity whose mean we can write as $\mu s + b$, where μ is a parameter that specifies the strength of the signal process. That is, the probability to find n events is,

$$P(n|\mu) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)}. \tag{17}$$

The values of $\mathbf{x}$ follow a pdf that is a mixture of the two contributions from signal and background,

$$f(\mathbf{x}|\mu) = \frac{\mu s}{\mu s + b} f(\mathbf{x}|s) + \frac{b}{\mu s + b} f(\mathbf{x}|b), \tag{18}$$

where the coefficients of each component give the fraction of events of each type.

The complete measurement thus consists of selecting n events and for each one measuring the two-dimensional vector quantity $\mathbf{x}$. The full likelihood is therefore,

$$L(\mu) = P(n|\mu) \prod_{i=1}^{n} f(\mathbf{x}_i|\mu) = \frac{e^{-(\mu s + b)}}{n!} \prod_{i=1}^{n} \left[\mu s f(\mathbf{x}_i|s) + b f(\mathbf{x}_i|b)\right]. \tag{19}$$

We can now carry out tests of different hypothetical values of μ. To establish the existence of the signal process we try to reject the hypothesis of the background-only model, $\mu = 0$. Regardless of whether we claim discovery, we can set limits on the signal strength μ, which we examine further in Sect. 11.

Let us first focus on the question of discovery, i.e., a test of $\mu = 0$. If the signal process exists, we would like to maximize the probability that we will discover it. This means that the test of the background-only ($\mu = 0$) hypothesis should have as high a power as possible relative to the alternative that includes signal ($\mu = 1$). According to the Neyman-Pearson lemma, the maximum power is achieved by basing the test on the likelihood ratio $L(1)/L(0)$, or equivalently on the statistic

$$Q = -2 \ln \frac{L(1)}{L(0)} = 2s - 2 \sum_{i=1}^{n} \ln \left(1 + \frac{s}{b} \frac{f(\mathbf{x}_i|s)}{f(\mathbf{x}_i|b)}\right). \tag{20}$$

The term $2s$ in front of the sum is a constant and so only shifts the distribution of Q for both hypotheses equally; it can therefore be dropped.

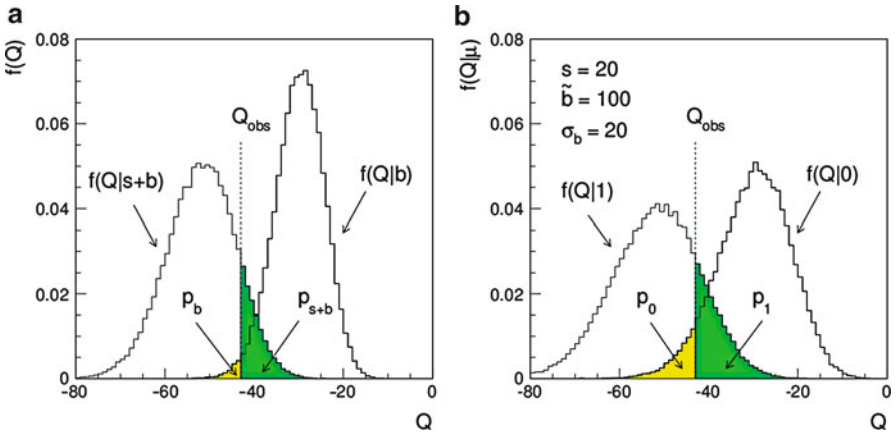

Fig. 6 (**a**) Distribution of the statistic Q assuming $s = 20$ and $b = 100$ under both the background-only ($\mu = 0$) signal-plus-background ($\mu = 1$) hypotheses; (**b**) same as in (**a**) but with b treated as having an uncertainty of $\sigma_b = 20$ (see text)

The other terms on the right-hand side of Eq. (20) are a sum of contributions from each event, and because the **x** values follow the same distribution for each event, each term in the sum follows the same distribution. To find the pdf of Q we can exploit the fact that the distribution of a sum of random variables is given by the convolution of their distributions. The full distribution can therefore be determined using Fourier transform techniques from the corresponding single-event distributions; details can be found in Ref. [18].

Following our toy example, suppose we take the expected numbers of events to be $b = 100$ for background and $s = 20$ for signal. The distribution of the statistic Q is found in this case simply by generating experiments according to the $\mu = 0$ and $\mu = 1$ hypotheses, computing Q for each according to Eq. (20) and recording the values in histograms. This results in the distributions shown in Fig. 6a.

To establish discovery of the signal process, we use the statistic Q to test the hypothesis that $\mu = 0$. As the test statistic is a monotonic function of the likelihood ratio $L(1)/L(0)$, we obtain maximum power relative to the alternative of $\mu = 1$. The p-value of $\mu = 0$ is computed as the area below Q_{obs} in Fig. 6, i.e. $p_0 = P(Q \leq Q_{\text{obs}}|0)$, because here lower Q corresponds to data more consistent with a positive μ (e.g., $\mu = 1$). The p-value can be converted into a significance Z using Eq. (9) and if this is greater than a specific threshold (e.g., 5.0) then one rejects the background-only hypothesis.

To set limits on μ we can use the statistic

$$Q_\mu = -2 \ln \frac{L(\mu)}{L(0)}, \tag{21}$$

defined such that the special case Q_1 is the same as the statistic Q used above for discovery. This will provide maximum power in a test of μ relative to the background-only alternative. The distribution of Q is also shown in Fig. 6 for $\mu = 1$. The p-value of the $\mu = 1$ hypothesis is given by the area above the observed value Q_{obs}, since higher values of Q are more consistent with the alternative of $\mu = 0$. This is shown here for the special case of $\mu = 1$ but one can repeat the procedure using $f(Q_\mu|\mu)$ for any other value of μ and compute the p-value p_μ in the analogous manner. To find the upper limit one would carry out the analysis as described above for all values of μ and reject those that have $p_\mu < \alpha$ for, say $\alpha = 0.05$. The highest value of μ not rejected is then the upper limit at 95 % C.L.

5.2 Systematic Uncertainties and Nuisance Parameters

Until now we have treated the expected number of background events b as known with negligible uncertainty. In practice, of course, this may not be true and so we may need to regard b as an adjustable parameter of our model. That is, we regard Eq. (19) as giving $L(\mu, b)$, where μ is the parameter of interest and b is a *nuisance parameter*.

There are several ways of eliminating the nuisance parameters from the problem. First we consider a method that is essentially frequentist but contains a Bayesian aspect. From the standpoint of nuisance parameters this is essentially the same as what is done in the purely Bayesian limits as discussed in Sect. 9. An alternative frequentist treatment using the profile likelihood is described in Sect. 6.

Consider first the frequentist method with Bayesian treatment of nuisance parameters. Our degree of belief about the true value of the parameter may be described in a Bayesian sense by a prior pdf $\pi(b)$ and our best estimate of b (e.g., the mean of $\pi(b)$) may be a value $\tilde{b}$. As an example, $\pi(b)$ could be a Gaussian distribution centred about $\tilde{b}$ with a standard deviation σ_b:

$$\pi(b) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(b-\tilde{b})^2/2\sigma_b^2}. \tag{22}$$

In fact a Gaussian pdf for b may not be the most appropriate model, e.g., if a parameter is bounded to be positive or if the prior should be characterized by longer positive tail. As an alternative one may use a Gaussian distribution for $\ln b$, which is to say that the pdf for b is log-normal.

Using the pdf $\pi(b)$ we can construct what is called the marginal (or prior predictive) likelihood,

$$L_{\text{m}}(n, \mathbf{x}_1, \ldots, \mathbf{x}_n|\mu) = \int L(n, \mathbf{x}_1, \ldots, \mathbf{x}_n|\mu, b)\pi(b)\, db, \tag{23}$$

where in the notation above we have emphasized that the likelihood of a model is the probability for the data under assumption of that model.

Notice that the marginal model does not represent the probability of data that would be generated if we were really to repeat the experiment. In that case we would not know the true value of b, but we could at least assume it would not change under repetition of the experiment. Rather, the marginal model represents a situation in which every repetition of the experiment is carried out with a new value of b randomly sampled from $\pi(b)$. It is in effect an average of models each with a given b, where the average is carried out with respect to the density $\pi(b)$.

For our tests we can use the same test statistic Q as before, but now we need to know its distribution under assumption of the prior predictive model. That is, if b is known exactly then we obtain distributions $f(Q|\mu, b)$ such as those shown in Fig. 6a. What we want instead is the distribution based on data that follows the marginal model,

$$f_{\mathrm{m}}(Q|\mu) = \int f(Q|\mu, b)\pi(b)\, db. \tag{24}$$

Although it may not be obvious how to compute this integral, it can be done easily with Monte Carlo by generating a value of b according to $\pi(b)$, then using this value to generate the data n, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and with these we find a value of Q which is recorded in a histogram. By repeating the entire procedure a large number of times we obtain distributions as shown in Fig. 6b, which are generated with a Gaussian prior for b with $\tilde{b} = 100$ and $\sigma_b = 20$.

As can be seen in Fig. 6, the effect of the uncertainty on b broadens the distributions of Q such that the p-values for both hypotheses are increased. That is, one may be able to reject one or the other hypothesis in the case where b was known because the p-value may be found less than α. When the uncertainty in b is included, however, the p-values may no longer allow one to reject the model in question.

As a further step one could consider using the marginal likelihood as the basis of the likelihood ratio used in the test statistic, i.e., we take $Q = -2\ln(L_{\mathrm{m}}(1)/L_{\mathrm{m}}(0))$. Use of a different statistic simply changes the critical region of the test and thus alters the power relative to the alternative models considered. This step by itself, however, does not take into account the uncertainty in b and it will not result in a broadening of $f(Q|\mu)$ and an increase in p-values as illustrated above. This is achieved by generating the distribution of Q using the marginal model through Eq. (24). In practice the marginal likelihoods can be very difficult to compute and a test statistic based on their ratio is not often used in HEP (see, however, Ref. [19]).

The ratio of marginal likelihoods is also called the *Bayes factor*, usually written with indices to denote the hypotheses being compared, e.g.,

$$B_{10} = \frac{L_{\mathrm{m}}(1)}{L_{\mathrm{m}}(0)}. \tag{25}$$

This is by itself a quantity of interest in Bayesian statistics as it represents the ratio of posterior probabilities of the hypotheses $\mu = 1$ and $\mu = 0$ in the special case where the prior probabilities are taken equal. If the Bayes factor is greater than one

it means that the evidence from the data results in an increase in one's belief in the hypothesis $\mu = 1$ over $\mu = 0$. Further discussion on the use of Bayes factors can be found in Refs. [1, 20].

Another possibility is to construct the test statistic from the ratio of *profile likelihoods*. Suppose the likelihood depends on a parameter of interest μ and nuisance parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. The profile likelihood L_p is defined as,

$$L_\mathrm{p}(\mu) = L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu)), \tag{26}$$

where $\hat{\hat{\boldsymbol{\theta}}}(\mu)$, called the profiled values of the nuisance parameters $\boldsymbol{\theta}$, are the values that maximizes $L(\mu, \theta)$ for the specified value of μ. Thus the profile likelihood only depends on μ. Searches at the Tevatron (e.g., Ref. [21]) have used the statistic,

$$Q = -2 \ln \frac{L_\mathrm{p}(1)}{L_\mathrm{p}(0)}. \tag{27}$$

As mentioned above, use of this statistic does not in itself take into account the systematic uncertainties related to the nuisance parameters. In Ref. [21] this has been done by generating the distribution of Q using the marginal model (23). An alternative to this procedure is to construct the statistic from a different profile likelihood ratio as described in Sect. 6.

6 Tests Based on the Profile Likelihood Ratio

Suppose as before that the parameter of interest is μ and the problem may contain one or more nuisance parameters $\boldsymbol{\theta}$. An alternative way to test hypothetical values of μ is to use the *profile likelihood ratio*,

$$\lambda(\mu) = \frac{L_\mathrm{p}(\mu)}{L(\hat{\mu}, \hat{\theta})}, \tag{28}$$

where L_p is the profile likelihood defined in Eq. (26) and $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximize the likelihood. In some models it may be that μ can only take on values in a restricted range, e.g., $\mu \geq 0$ if this parameter is proportional to the cross section of the signal process. In this case we can, however, regard $\hat{\mu}$ as an effective estimator that is allowed to take on negative values. This will allow us to write down simple formulae for the distributions of test statistics that are valid in the limit where the data sample is very large.

The quantity $\lambda(\mu)$ is defined so that it lies between zero and one, with higher values indicating greater compatibility between the data and the hypothesized value of μ. We can therefore use $\lambda(\mu)$ to construct a statistic to test different values of μ. Suppose as above that μ is proportional to the rate of the sought after signal process and we want to test the background-only ($\mu = 0$) hypothesis.

Often the signal process is such that only positive values of μ are regarded as relevant alternatives. In this case we would choose the critical region of our test of $\mu = 0$ to correspond to data outcomes characteristic of positive μ, that is, when $\hat{\mu} > 0$. It could happen that we find $\hat{\mu} < 0$, e.g., if the total observed number of events fluctuates below what is expected from background alone. Although a negative $\hat{\mu}$ indicates a level of incompatibility between the data and hypothesis of $\mu = 0$, this is not the type of disagreement that we want to exploit to declare discovery of a positive signal process.

Providing our signal models are of the type described above, we can take the statistic used to test $\mu = 0$ to be,

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{29}$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$ as defined in Eq. (28). In this way, higher values of q_0 correspond to increasing disagreement between data and hypothesis, and so the p-value of $\mu = 0$ is the probability, assuming $\mu = 0$ to find q_0 at least high or higher than the observed value.

If we are interested in an upper limit for the parameter μ, then we want the critical region to correspond to data values characteristic of the alternative $\mu = 0$. This can be achieved by defining,

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{30}$$

For both discovery and upper limits, therefore, the p-value for a hypothesized μ is then,

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu, \theta)\, dq_\mu. \tag{31}$$

If we use the statistic q_μ then we find the upper limit μ_{up} at confidence level $1-\alpha$ by setting $p_\mu = \alpha$ and solving for μ. This will have the property $P(\mu_{\mathrm{up}} \geq \mu) \geq 1-\alpha$. Note that the p-value pertains to the hypothesis of not only μ but also the nuisance parameters $\boldsymbol{\theta}$. We will return to this point below.

To find the p-value we need the distribution of the test statistic under assumption of the same μ being tested. For sufficiently large data samples one can show that this distribution approaches an asymptotic form related to the chi-square distribution, where the number of degrees of freedom is equal to the number of parameters of interest (in this example just one, i.e., μ). The asymptotic formulae are based on theorems due to Wilks [22] and Wald [23] and are described in further detail in Ref. [24].

An important advantage of using the profile likelihood ratio is that its asymptotic distribution is independent of the nuisance parameters, so we are not required to choose specific values for them to compute the p-value. In practice one has of course a finite data sample and so the asymptotic formulae are not exact. Therefore the p-values will in general depend on the nuisance parameters to some extent.

Providing the conditions for the asymptotic approximations hold, one finds a very simple formula for the p-value,

$$p_\mu = \Phi\left(\sqrt{q_\mu}\right), \tag{32}$$

where Φ is the cumulative distribution of the standard Gaussian. From Eq. (9) we find for the corresponding significance,

$$Z_\mu = \sqrt{q_\mu}. \tag{33}$$

For discovery, we could require Z_0 greater than some threshold such as 5.0, which corresponds to $p_0 < 2.9 \times 10^{-7}$. When setting limits one usually excludes a parameter value if its p-value is less than, say, 0.05, corresponding to a confidence level of 95 %, or a significance of 1.64. Although Eqs. (32) and (33) are only exact for an infinitely large data sample, the approach to the asymptotic limit is very fast and the approximations often turn out to be valid for moderate or even surprisingly small data samples. Examples can be found in Ref. [24].

For data samples not large enough to allow use of the asymptotic formulae, one must determine the distribution of the test statistics by other means, e.g., with Monte Carlo models that use specific values for the nuisance parameters. In the exact frequentist approach we would then only reject a value of μ if we find its p-value less than α for all possible value of the nuisance parameters. Therefore only a smaller set of μ values are rejected and the resulting confidence interval becomes larger, which is to say the limits on μ become less stringent. The confidence interval then *overcovers*, i.e., its probability to contain the true μ is greater than $1 - \alpha$, at least for some values of the nuisance parameters.

It may seem unfortunate if we cannot reject values of μ that are retained only under assumption of nuisance parameter values that may be highly disfavoured, e.g., for theoretical reasons. A compromise solution is test μ using the p-value based only on the profiled values of the nuisance parameters, i.e., we take

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu, \hat{\hat{\theta}}(\mu))\, dq_\mu. \tag{34}$$

This procedure has been called *profile construction* [25] in HEP or *hybrid resampling* [26,27] amongst statisticians. If the true values of the nuisance parameters are equal to the profiled values, then the coverage probability of the resulting confidence interval for μ is exact. For other values of θ, the interval for μ may over- or undercover. In cases where this is crucial, one may include a wider range of nuisance parameter values and study the coverage with Monte Carlo.

7 Summary on Likelihood Ratios

Above we have seen two closely related ways to construct tests based on likelihood ratios and also two different ways of incorporating systematic uncertainties. In Sect. 5 we used the ratio of two simple hypotheses, namely $L(\mu)/L(0)$, whereas in Sect. 6 the statistic used was $L(\mu)/L(\hat{\mu})$.

If there are no nuisance parameters in the model, then the Neyman-Pearson lemma guarantees that the ratio $L(\mu)/L(0)$ provides the greatest power in a test of $\mu = 0$ with respect to the alternative of μ. If there are nuisance parameters then this will not in general hold. In this case one can replace the likelihood L by the marginal or profile likelihood, which results in a different critical region for the test. It can be difficult to find the exact power for different alternatives but one can study this using Monte Carlo. The important point is that by changing the critical region of the test by using a ratio of marginal or profile likelihoods, one does not (by this step alone) account for the systematic uncertainties.

To include the uncertainties reflected by the nuisance parameters into the test we have also seen two approaches. One has been to construct the marginal (prior predictive) model (23) to determine the distribution of the test statistic. If, for example, one rejects the hypothesis $\mu = 0$, then the model rejected represents an average of models corresponding to different values of the nuisance parameters. This is the approach we used together with the likelihood ratio $L(\mu)/L(0)$ (with or without the marginal or profile likelihoods in the ratio).

In contrast to this, when we used the statistic based on the profile likelihood ratio $L(\mu, \hat{\hat{\boldsymbol{\theta}}})/L(\hat{\mu}, \hat{\boldsymbol{\theta}})$ we exploited the fact that its distribution becomes independent of the nuisance parameters in the large sample limit. In this case we are able to say that if $\mu = 0$ is rejected, then this holds for all values of the nuisance parameters $\boldsymbol{\theta}$. The large-sample distributions based on Wilks' theorem are only valid when the likelihood ratio is constructed in this way; this is not the case, e.g., if one were to characterize nuisance parameters with a prior and then marginalize (integrate).

In a real analysis the data sample is finite and so the p-values for the parameter of interest μ will depend at some level on the nuisance parameters $\boldsymbol{\theta}$. In such a case one may then use their profiled values $\hat{\hat{\boldsymbol{\theta}}}$ under assumption of the value of μ being tested. If these are equal to the true values of the nuisance parameters, then the p-values for μ will be correct and a confidence interval for μ will cover the true value with a probability equal to the nominal confidence level $1 - \alpha$. If the true values of $\boldsymbol{\theta}$ are not equal to the profiled values, then the p-values may be too high or to low, which is to say that confidence intervals for μ may be to large or too small.

Both types of likelihood ratios and more importantly, both methods for determining their sampling distributions (averaged or not) are widely used. For many analyses they will lead to very similar conclusions. An important advantage of the profile likelihood ratio is that one can say what set of physical models have been rejected (i.e., what points in nuisance parameter space). If necessary, Monte Carlo studies can be carried out to obtain the p-values using nuisance parameters in some region about their profiled values $\hat{\hat{\boldsymbol{\theta}}}(\mu)$.

8 Unified Intervals

The test of μ used for an upper limit assumes that the relevant alternative hypothesis is $\mu = 0$, and the critical region is chosen accordingly. In other cases, one may regard values of μ both higher and lower than the one being tested as valid alternatives, and one would therefore like a test that has high power for both cases. One can show that in general there is no single test (i.e., no given critical region) that will have the highest power relative to all alternatives (see, e.g., Ref. [9], Chapter 22).

Nevertheless, we can use the statistic

$$t_\mu = -2 \ln \lambda(\mu), \tag{35}$$

to construct a test for any value of μ. As before, higher values of the statistic correspond to increasing disagreement between the data and the hypothesized μ. Here, however, the critical region can include data corresponding to an estimated signal strength $\hat{\mu}$ greater or less than μ. If one carries out a test of all values of μ using this statistic, then both high and low values of μ may be rejected.

Suppose the lowest and highest values not rejected are μ_1 and μ_2, respectively. One may be tempted to interpret the upper edge of such an interval as an upper limit in the same sense as the one derived above using q_μ from Eq. (30). The coverage probability, however, refers to the whole interval, i.e., one has $P(\mu_1 \leq \mu \leq \mu_2) \geq 1 - \alpha$. One cannot in general make a corresponding statement about the probability for the upper or lower edge of the interval alone to be above or below μ, analogous to the statement $P(\mu_{\text{up}} \geq \mu) \geq 1 - \alpha$ that holds for an upper limit.

The confidence intervals proposed by Feldman and Cousins [28], also called *unified intervals*, are based on a statistic similar to t_μ from Eq. (35) with the additional restriction that the estimator $\hat{\mu}$ that appears in the denominator of the likelihood ratio is restricted to physically allowed values of μ. Large-sample formulae for the distributions and corresponding p-values can be found in Ref. [24]. (In that reference the statistic for the case $\mu \geq 0$ is called $\tilde{t}_\mu$.) The problem of excluding parameter values for which one has no sensitivity is mitigated with unified intervals by the particular choice of the critical region of the test (see Ref. [28]).

9 Bayesian Limits

Although these lectures focus mainly on frequentist statistical procedures we provide here a brief description of the Bayesian approach to setting limits. This is in fact conceptually much simpler than the frequentist procedure. Suppose we have a model that contains a parameter μ, which, as before, we imagine as being proportional to the rate of a sought-after signal process. In addition, the model may contain some nuisance parameters θ. As in the frequentist case, we will have a

likelihood $L(\mathbf{x}|\mu, \theta)$ which gives the probability for the data $\mathbf{x}$ given μ and θ. In a Bayesian analysis we are allowed to associate a probability with parameter values, and so we assess our degree of belief in a given model (or set of parameter values) by giving the posterior probability $p(\mu, \theta|\mathbf{x})$. To find this we use Bayes' theorem (4), which we can write as a proportionality,

$$p(\mu, \theta|\mathbf{x}) \propto L(\mathbf{x}|\mu, \theta)\pi(\mu, \theta), \tag{36}$$

where the prior pdf $\pi(\mu, \theta)$ specifies our degree of belief in the parameters' values before carrying out the measurement.

The problematic ingredient in the procedure above is the prior pdf $\pi(\mu, \theta)$. For a nuisance parameter θ, one typically has some specific information that constrains one's degree of belief about its value. For example, a calibration constant or background event rate may be constrained by some control measurements, leading to a best estimate $\tilde{\theta}$ and some measure of its uncertainty σ_θ. Depending on the problem at hand one may from these subsidiary measurements as well as physical or theoretical constraints construct a prior pdf for θ. In many cases this will be independent of the value of the parameter of interest μ, in which case the prior will factorize, i.e., $\pi(\mu, \theta) = \pi_\mu(\mu)\pi_\theta(\theta)$. For the present discussion we will assume that this is the case.

The more controversial part of the procedure is the prior $\pi_\mu(\mu)$ for the parameter of interest. As one is carrying out the measurement in order to learn about μ, one usually does not have much information about it beforehand, at least not much relative to the amount one hopes to gain. Therefore one may like to write down a prior that is *non-informative*, i.e., it reflects a maximal degree of prior ignorance about μ, in the hopes that one will in this way avoid injecting any bias into the result. This turns out to be impossible, or at least there is no unique way of quantifying prior ignorance.

As a first attempt at a non-informative prior for μ we might choose to take it very broad relative to the likelihood. Suppose as before that μ represents the rate of signal so we have $\mu \geq 0$. As an extreme example of a broad prior we may try,

$$\pi_\mu(\mu) = \begin{cases} 1 & \mu \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

This so-called flat prior is problematic for a number of reasons. First, it cannot be normalized to unit area, so it is not a proper pdf; it is said to be *improper*. Here this defect is not fatal because in Bayes' theorem the prior always appears multiplied by the likelihood, and if this falls off sufficiently rapidly as a function of μ, as is often the case in practice, then the posterior pdf for μ may indeed be normalizable.

A further difficulty with a flat prior is that our inference is not invariant under a change in parameter. For example, if we were to take as the parameter $\eta = \ln \mu$, then according to the rules for transformation of variables we find for the pdf of η,

$$\pi_\eta(\eta) = \pi_\mu(\mu) \left| \frac{d\mu}{d\eta} \right| = e^\eta \pi_\mu(\mu(\eta)), \tag{38}$$

so if $\pi_\mu(\mu)$ is constant then $\pi_\eta(\eta) \propto e^\eta$ which is not. So if we claim we know nothing about μ and hence use for it a constant prior, we are implicitly saying that we known something about η.

Finally we should note that the constant prior of Eq. (37) cannot in any realistic sense reflect a degree of belief, since it assigns a zero probability to the range between any two finite limits.

The difficult and subjective nature of encoding personal knowledge into priors has led to what is called *objective Bayesian statistics*, where prior probabilities are based not on an actual degree of belief but rather derived from formal rules. These give, for example, priors which are invariant under a transformation of parameters or which result in a maximum gain in information for a given set of measurements. For an extensive review see, for example, Ref. [29]; applications to HEP are discussed in Refs. [30, 31].

The constant prior of Eq. (37) has been used in HEP so widely that it serves a useful purpose as a benchmark, despite its shortcomings. Although interpretation of the posterior probability as a degree of belief is no longer strictly true, one can simply regard the resulting interval as a given function of the data, which will with some probability contain the true value of the parameter. Unlike the confidence interval obtained from the frequentist procedure, however, the coverage probability will depend in general on the true (and unknown) value of the parameter.

We now turn to the Bayesian treatment of nuisance parameters. What we get from Bayes' theorem is the joint distribution of all of the parameters in the problem, in this case both μ and θ. Because we are not interested in the nuisance parameter θ we simply integrate (or sum in the case of a discrete parameter) to find the marginal pdf for the parameter of interest, i.e.,

$$p(\mu|\mathbf{x}) = \int p(\mu, \theta|\mathbf{x})\, d\theta. \tag{39}$$

One typically has not one but many nuisance parameters and the integral required to marginalize over them cannot be carried out in closed form. Even Monte Carlo integration based on the acceptance-rejection method becomes impractical if the number of parameters is too large, since then the acceptance rate becomes very small. In such cases, Markov Chain Monte Carlo (MCMC) provides an effective means to calculate integrals of this type. Here one generates a correlated sequence of points in the full parameter space and records the distribution of the parameter of interest, in effect determining its marginal distribution. An MCMC method widely applicable to this sort of problem is the Metropolis-Hastings algorithm, which is described briefly in Ref. [14]. In-depth treatments of MCMC can be found, for example, in the texts by Robert and Casella [32], Liu [33], and the review by Neal [34].

10 The Poisson Counting Experiment

As a simple example, consider an experiment in which one counts a number of events n, modeled as following a Poisson distribution with a mean of $s + b$, where s and b are the contributions from signal and background processes, respectively. Suppose that b is known and we want to test different hypothetical values of s. Specifically, we want to test the hypothesis of $s = 0$ to see if we can establish the existence of the signal, and regardless of whether we succeed in doing this, we can set an upper limit on s.

To establish discovery of the signal using the frequentist approach, we test $s = 0$ against the alternative of $s > 0$. That is, we assume the relevant signal models imply positive s, and therefore we take the critical region of the test to correspond to larger numbers of events. Equivalently, we can define the p-value of the $s = 0$ hypothesis to be the probability, assuming $s = 0$, to find as many events as actually observed or more, i.e.,

$$p_0 = P(n \geq n_{\text{obs}} | s = 0, b) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b}. \tag{40}$$

We can exploit a mathematical identity,

$$\sum_{n=0}^{m} P(n|b) = 1 - F_{\chi^2}(2b; n_{\text{dof}}), \tag{41}$$

with $n_{\text{dof}} = 2(m + 1)$ to relate the sum of Poisson probabilities in Eq. (40) to the cumulative chi-square distribution F_{χ^2}, which allows us write the p-value as,

$$p_0 = F_{\chi^2}(2b; 2n_{\text{obs}}). \tag{42}$$

For example, suppose $b = 3.4$ and we observe $n_{\text{obs}} = 16$ events. Equation (42) gives $p_0 = 3.6 \times 10^{-6}$ corresponding to a significance $Z = 4.5$. This would thus constitute strong evidence in favour of a nonzero value of s, but is still below the traditional threshold of $Z = 5$.

To construct the frequentist upper limit, we should test all hypothetical values of s against the alternative of $s = 0$, so the critical region consists of low values of n. This means we take the p-value of a hypothesized s to be the probability to find n as small as observed or smaller, i.e.,

$$p_s = \sum_{m=0}^{n} \frac{(s + b)^m}{m!} e^{-(s+b)}. \tag{43}$$

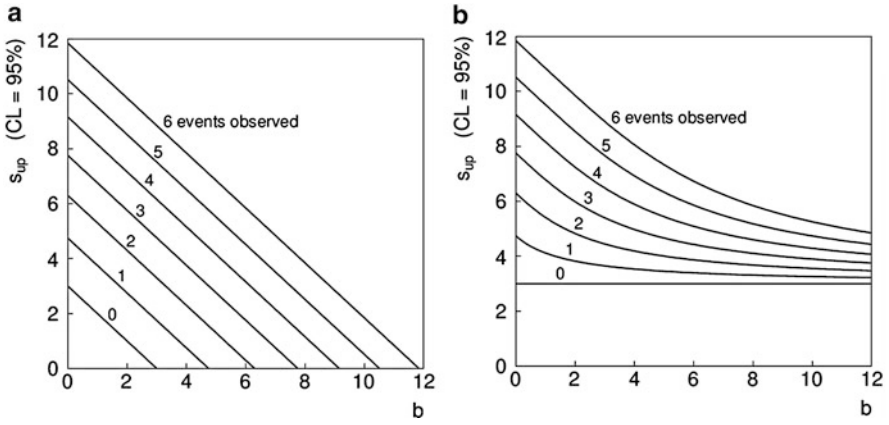

Fig. 7 Upper limits on the mean number of signal events s at 95 % confidence level as a function of the expected background b for (**a**) the frequentist method and (**b**) Bayesian method with a flat prior

The upper limit at CL $= 1 - \alpha$ is found from the value of s such that the p-value is equal to α, i.e.,

$$\alpha = \sum_{m=0}^{n} \frac{(s_{\text{up}} + b)^m}{m!} e^{-(s_{\text{up}}+b)} = 1 - F_{\chi^2}\left[2(s_{\text{up}} + b), 2(n+1)\right], \tag{44}$$

where in the second equality we again used the identity (41) to relate the sum of Poisson probabilities to the cumulative chi-square distribution. This allows us to solve for the upper limit,

$$s_{\text{up}} = \frac{1}{2} F_{\chi^2}^{-1}\left[1 - \alpha, 2(n+1)\right] - b, \tag{45}$$

where $F_{\chi^2}^{-1}$ is the chi-square quantile (inverse of the cumulative distribution). The upper limit s_{up} is shown in Fig. 7a for $1 - \alpha = 95\,\%$ as a function of b for different numbers of observed events n.

To find the corresponding upper limit in the Bayesian approach, we need to assume a prior pdf for s. If we use the flat prior of Eq. (37), then by using Bayes' theorem we find the posterior pdf,

$$p(s|n) \propto \frac{(s+b)^n}{n!} e^{-(s+b)}, \tag{46}$$

for $s \geq 0$ and $p(\mu|n) = 0$ otherwise. This can be normalized to unit area, which gives,

$$p(s|n) = \frac{(s+b)^n e^{-(s+b)}}{\Gamma(b, n+1)}, \tag{47}$$

where $\Gamma(b, n+1) = \int_b^\infty x^n e^{-x}\, dx$ is the upper incomplete gamma function.

Since in the Bayesian approach we are assigning a probability to s, we can express an upper limit simply by integrating the posterior pdf from the minimum value $s = 0$ up to an upper limit s_{up} such that this contains a fixed probability, say, $1 - \alpha$. That is, we require,

$$1 - \alpha = \int_0^{s_{\text{up}}} p(s|n)\, ds. \tag{48}$$

To solve for s_{up}, we can use the integral

$$\int_0^a x^n e^{-x}\, dx = \Gamma(n+1) F_{\chi^2}(2a, 2(n+1)), \tag{49}$$

where again F_{χ^2} is the cumulative chi-square distribution for $2(n+1)$ degrees of freedom. Using this we find for the upper limit

$$s_{\text{up}} = \frac{1}{2} F_{\chi^2}^{-1}\left[p, 2(n+1)\right] - b, \tag{50}$$

where,

$$p = 1 - \alpha \left(1 - F_{\chi^2}\left[2b, 2(n+1)\right]\right). \tag{51}$$

This is shown in Fig. 7b. Interestingly, the upper limits for the case of $b = 0$ happen to coincide exactly with the values we found for the frequentist upper limit, and for nonzero b the Bayesian limits are everywhere higher. This means that the probability for the Bayesian interval to include the true value of s is higher than $1 - \alpha$, so in this sense one can say that the Bayesian limit is conservative. The corresponding unified interval from the procedure of Feldman-Cousins is described in Ref. [28].

If the parameter b is not known, then this can be included in the limit using the methods discussed above. That is, one must treat b as a nuisance parameter, and in general one would have some control measurement that constrains its value. In the frequentist approach b is eliminated by profiling; in the Bayesian case one requires a prior pdf for b and simply marginalizes the joint pdf of s and b to find the posterior $p(s|n)$. The problem of a Poisson counting experiment with additional nuisance parameters is discussed in detail in Refs. [30, 35, 36].

11 Limits in Cases of Low Sensitivity

An important issue arises when setting frequentist limits that is already apparent in the example from Sect. 10. In Fig. 7a, which shows the frequentist upper limit on the parameter s as a function of b, one sees that $s_{\rm up}$ can be arbitrarily small. Naive application of Eq. (45) can in fact result in a negative upper limit for what should be an intrinsically positive quantity. What this really means that all values of s are rejected in a test of size α. This can happen if the number of observed events n fluctuates substantially below the expected background b. One is then faced with the prospect of not obtaining a useful upper limit as the outcome of one's expensive experiment. It might be hoped that such an occurrence would be rare but by construction it should happen with probability α, e.g., 5 % of the time.

Essentially the same problem comes up whenever we test any hypothesis to which we have very low sensitivity. What "low sensitivity" means here is that the distributions of whatever statistic we are using is almost the same under assumption of the signal model being tested as it is under the background-only hypothesis. This type of situation is illustrated in Fig. 8a, where here we have labeled the model including signal $s + b$ (in our previous notation, $\mu = 1$) and the background-only model b (i.e., $\mu = 0$).

The critical region for a test of the $s + b$ hypothesis consists of high values of Q. Equivalently, the p-value is the probability $p_{s+b} = P(Q \geq Q_{\rm obs}|s + b)$. Because the distributions of Q under both hypotheses are very close, the power of the test of $s + b$ is only slightly greater than the size of the test α, which is equivalent to the statement that the quantity $1 - p_b$ is only slightly greater than p_{s+b}.

If we have no sensitivity to a particular model, such as the hypothesis of a Higgs boson with a mass much greater than what we could produce in our experiment, then we do not want to reject it, since our measurement can produce no evidence to

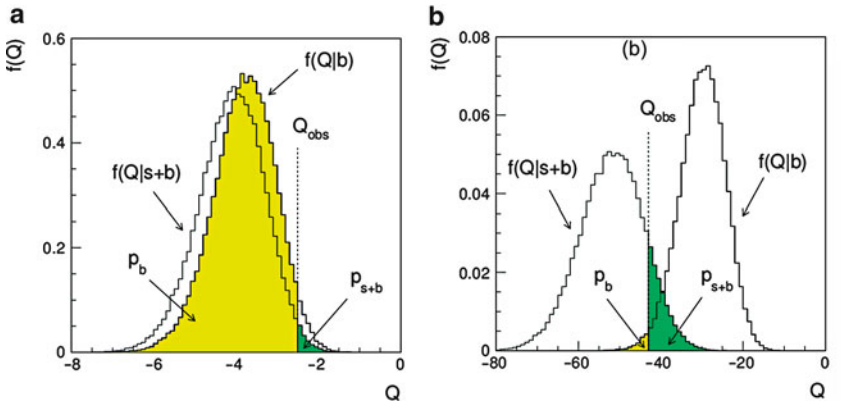

Fig. 8 (**a**) Distributions of the statistic Q indicating low sensitivity to the hypothesized signal model; (**b**) illustration of the ingredients for the CL_s limit

justify such a claim. Unfortunately, the frequentist procedure that rejects the signal model if its p-value is found less then α will do just that with a probability of at least α. And this will happen even if the model is, from an experimental standpoint, virtually indistinguishable from the background-only hypothesis. Since we typically take $\alpha = 0.05$, we will exclude one model out of every twenty to which we have no sensitivity.

One solution to this problem is the CL_s procedure proposed by Alex Read [37, 38], whereby the threshold for rejecting a model is altered in a way that prevents one from rejecting a model in the limit that one has very little sensitivity, but reverts to the usual frequentist procedure when the sensitivity is high. This is achieved by defining,

$$\mathrm{CL}_s = \frac{P(Q \geq Q_{\mathrm{obs}}|s+b)}{P(Q \geq Q_{\mathrm{obs}}|b)} = \frac{p_{s+b}}{1-p_b}. \tag{52}$$

The quantity CL_s then is then used in place of the p-value p_{s+b}, i.e., the $s+b$ model is rejected if one finds $\mathrm{CL}_s \leq \alpha$. The ingredients are illustrated in Fig. 8b.

One can understand qualitatively how this achieves the desired goal by considering the case where the distributions of Q under the two hypotheses $s+b$ and b are close together. Suppose the observed value Q_{obs} is such that p_{s+b} is less than α, so that in the usual frequentist procedure we would reject the $s+b$ hypothesis. In the case of low sensitivity, however, the quantity $1-p_b$ will also be small, as can be seen from Fig. 8a. Therefore, the quantity CL_s will be greater than p_{s+b} such that the $s+b$ model is not rejected by the criterion of Eq. (52).

If, on the other hand, the distributions are well separated, and Q_{obs} is such that the $p_{s+b} < \alpha$, then p_b will also be small and the term $1-p_b$ that appears in the denominator of CL_s will be close to unity. Therefore, in the case with high sensitivity, using CL_s is similar to what is obtained from the usual frequentist procedure based on the p-value p_{s+b}.

The largest value of s not rejected by the CL_s criterion gives the corresponding CL_s upper limit. Here to follow the traditional notation we have described it in terms of the mean number of signal events s rather than the strength parameter μ, but it is equivalent to using $\mathrm{CL}_\mu = p_\mu/(1-p_0)$ to find an interval for μ.

The CL_s procedure described above assumes that the test statistic Q is continuous. The recipe is slightly different if the data are discrete, such as a Poisson distributed number of events n with a mean $s+b$. In this case the quantity CL_s is defined as,

$$\mathrm{CL}_s = \frac{P(n \leq n_{\mathrm{obs}}|s+b)}{P(n \leq n_{\mathrm{obs}}|b)}, \tag{53}$$

where n_{obs} is the number of events observed. Here the numerator is p_{s+b}, which the same as in Eq. (52). The p-value of the background-only hypothesis is $p_b = P(n \geq n_{\mathrm{obs}}|b)$, but the denominator in Eq. (53) requires n less than *or equal* to n_{obs}, so this

is not exactly the same as $1 - p_b$. Equation (53) is the fundamental definition and it reduces to the ratio of p-values for the case of a continuous test statistic.

For a Poisson distributed number of events, the CL_s upper limit coincides exactly with the Bayesian upper limit based on the flat prior as shown in Fig. 7b. It is thus also greater than or equal to the limit based on the p-value and is in this sense conservative. It also turns out that the CL_s and Bayesian limits (using a flat prior) agree for the important case of Gaussian distributed data [37]. The problem of exclusion in the case of little or no sensitivity is mitigated in a different way by the unified intervals seen in Sect. 8 by the particular choice of the critical region (see, e.g., Ref. [28]).

12 The Look-Elsewhere Effect

Recently there has been important progress made on the problem of multiple testing, usually called in particle physics the "look-elsewhere effect" [39, 40]. The problem often relates to finding a peak in a distribution when the peak's position is not predicted in advance. In the frequentist approach, the correct p-value of the no-peak hypothesis is the probability, assuming background only, to find a peak as significant as the one found anywhere in the search region. This can be substantially higher than the probability to find a peak of equal or greater significance in the particular place where it appeared.

The "brute-force" solution to this problem involves generating data under the background-only hypothesis and for each data set, fitting a peak of unknown position and recording a measure of its significance. To establish a discovery, one often requires a p-value less than 2.9×10^{-7}, corresponding to a 5σ effect. Thus determining this with Monte Carlo requires generating and fitting an enormous number of experiments, perhaps several times 10^7. This is particularly difficult in that under the background-only hypothesis there is no real peak, but only fluctuations. One of these fluctuations will stand out as the most significant peak and this must be found in order to determine the value of the test statistic such as the profile likelihood ratio, $L(0)/L(\hat{\mu})$, for that particular data set. This must be repeated tens of millions of times without failure of the fitting program, which is a difficult computational challenge.

In contrast, if the position of the peak were known in advance, then the fit to the distribution would be much faster and easier, and furthermore, one can in many cases use formulae valid for sufficiently large samples that bypass completely the need for Monte Carlo (see, e.g., [24]). But this "fixed-position" p-value would not be correct in general, as it assumes the position of the peak was known in advance.

Gross and Vitells [39] have described a method that allows one to modify the p-value computed under assumption of a fixed position to obtain the correct value using a relatively simple calculation. Suppose a test statistic q_0, defined so that larger values indicate increasing disagreement with the data, is observed to have a value u. Furthermore, suppose the model contains a nuisance parameter θ (such as

the peak position) which is only defined under the signal model (there is no peak in the background-only model). An approximation for the desired "global" p-value is found to be,

$$p_{\text{global}} \approx p_{\text{local}} + \langle N_u \rangle, \tag{54}$$

where p_{local} is the p-value assuming a fixed value of θ (e.g., fixed peak position), and $\langle N_u \rangle$ is the mean number of "upcrossings" of the statistic q_0 above the level u in the range of the nuisance parameter considered (e.g., the mass range).

The value of $\langle N_u \rangle$ can be estimated from the number of upcrossings $\langle N_{u_0} \rangle$ above some much lower value, u_0, by using a relation due to Davis [41],

$$\langle N_u \rangle \approx \langle N_{u_0} \rangle e^{-(u-u_0)/2}. \tag{55}$$

By choosing u_0 sufficiently low, the value of $\langle N_u \rangle$ can be estimated by simulating only a very small number of experiments, rather than the 10^7 needed if one is dealing with a 5σ effect.

Vitells and Gross also indicate how to extend the correction to the case of more than one parameter, e.g., where one searches for a peak of both unknown position and width, or for searching for a peak in a two-dimensional space, such as an astrophysical measurement on the sky [40]. Here one may find some number of regions where signal appears to be present, but within those regions there may be islands or holes where the significance is lower. In the generalization to multiple dimensions, the number of upcrossings of the test statistic q_0 is replaced by the expectation of a quantity called the Euler characteristic, which is roughly speaking the number of disconnected regions with significant signal minus the number of 'holes'.

It should be emphasized that an exact accounting of the look-elsewhere effect requires that one specify where else one looked, e.g., the mass range in which a peak was sought. But this may be have been defined in a somewhat arbitrary manner, and one might have included not only the mass range but other variables that were also inspected for peaks but where none was found. It is therefore not worth expending great effort on an exact treatment of the look-elsewhere effect, as would be needed in the brute-force method mentioned above. Rather, the more easily obtained local p-value can be reported along with an approximate correction to account for the range of measurements in which the effect could have appeared.

13 Examples from the Higgs Search at the LHC

In this section we show how the methods described above have been applied to the recent discovery of a Higgs-like boson at the LHC. The examples are taken from the analyses of the ATLAS experiment [42]; similar results were obtained by CMS [43].

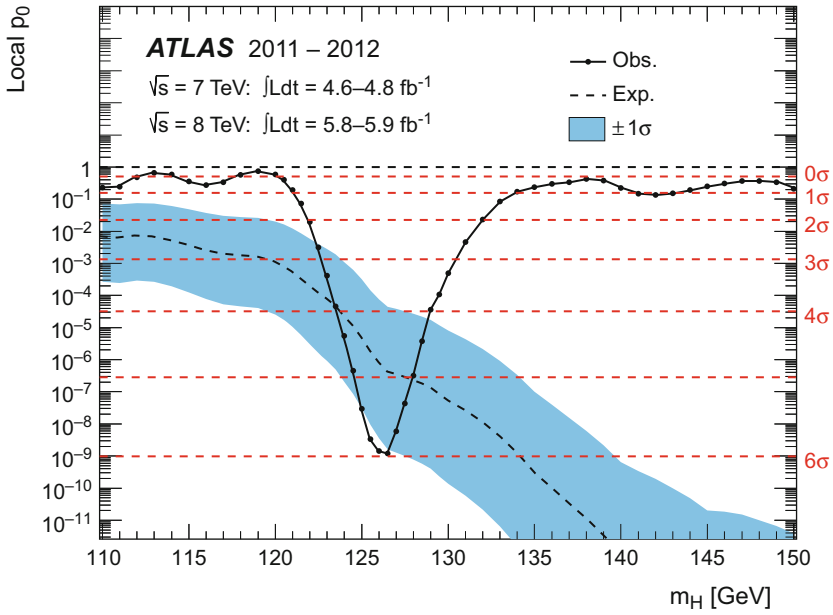

Fig. 9 The p-value of the background-only hypothesis versus the Higgs mass m_{H} (From Ref. [42]; see text)

The Higgs search is more complicated than examples described earlier because the production of something like a Higgs boson is characterized by two parameters of interest: the strength parameter μ, which is defined here as the signal cross section divided by the one predicted by the Standard Model, and the mass of the resonance, here labeled m_{H}. The procedure has been to carry out tests of μ for a set of fixed masses within a given range, and the results are then interpolated. One obtains from this two important outputs, both as a function of m_{H}: p-values for the test of $\mu = 0$ and confidence intervals (here, upper limits) for μ.

The p-value of the background-only hypothesis p_0 is shown versus m_{H} in Fig. 9. The values shown are not corrected for the look-elsewhere effect; this is therefore referred to as the *local* p_0. On the right-hand side of the plot one can see the value translated into the significance Z according to Eq. (9). The lowest p-value is found at $m_{\mathrm{H}} = 126.5\,\mathrm{GeV}$ and corresponds to $Z = 6.0$; taking into account some additional systematic uncertainties in the electromagnetic energy response reduces this to 5.9.

The correction for the look-elsewhere effect is based on the procedure described in Sect. 12 and in Ref. [39]. If the mass range of the search is taken to be 110–600 GeV, the peak significance Z reduces from 5.9 to 5.1; if one takes 110–150 GeV it gives $Z = 5.3$.

The dotted line in Fig. 9 gives the median value of Z under the hypothesis that the Higgs boson is present at the rate predicted by the Standard Model, i.e., $\mu = 1$.

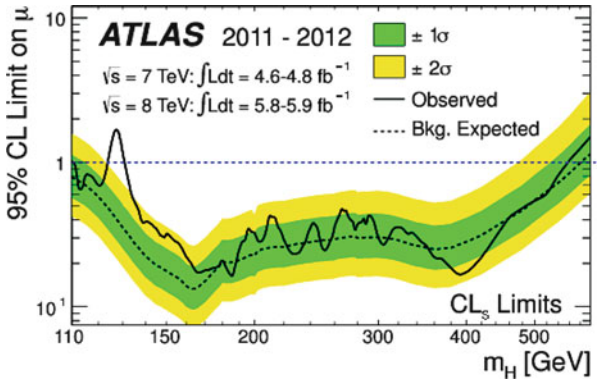

Fig. 10 CL_s upper limits on the production cross section for the Higgs boson as a function of its mass (From Ref. [42]; see text)

That is, if one were to generate a data set assuming an SM Higgs boson with a mass of 126.5 GeV, then this will lead to a certain significance Z for a test of $\mu = 0$. If one were to generate an ensemble of such experiments then the median of the resulting distribution of Z values, usually referred to as the expected significance, is taken as a measure of the sensitivity of the measurement. The median Z is preferred over the expectation value because this is still related to the median p-value through Eq. (9); for the expectation value the corresponding relation would not hold.

For $m_H = 126.5$ GeV, the expected significance is $Z = 4.9$, as can be seen from the dotted line. The blue band corresponds to the 68 % inter-quantile range, i.e., the lower and upper edges of the band are the 16 and 84 % quantiles of the distribution (referred to as the $\pm 1\sigma$ band). The band quantifies how much variation of the result to expect as a result of statistical fluctuations if the nominal signal model is correct. From Fig. 9 one can see that the observed p-value is at the lower edge of the blue band. So if the $\mu = 1$ hypothesis is in fact correct, then the signal rate observed by ATLAS fluctuated above the median value by a bit more than one standard deviation.

Figure 10 shows the upper limit on the strength parameter μ as a function of the Higgs mass. As with the case of $\mu = 0$ described above, the test procedure was carried out for a set of discrete values of the mass and the results interpolated. The solid curve shows the observed upper limit using the CL_s procedure described in Sect. 11. For each mass, the distribution of upper limits was found under assumption of background only, and the dotted curve shows the median value. The green and yellow bands show the 68 and 95 % inter-quantile ranges, i.e., the ranges that would correspond to $\pm 1\sigma$ and $\pm 2\sigma$ if the distribution were Gaussian. In fact, because the CL_s procedure prevents one from excluding very low values of μ, the distribution of upper limits can be significantly more asymmetric than a Gaussian.

For almost all mass values, the observed limit is close to the expectation under assumption of $\mu = 0$. The exception is the mass region around 126 GeV, where the upper limit is significantly higher. This of course corresponds to the discovered signal.

14 Why 5 σ?

Common practice in HEP has been to regard an observed signal to be worthy of the word "discovery" when its significance exceeds $Z = 5$, corresponding to a p-value of the background-only hypothesis of 2.9×10^{-7}. This is in stark contrast to many other fields (e.g., medicine, psychology) in which a p-value of 5 % ($Z = 1.64$) is considered significant. In this section, we examine critically some of the reasons why the community has used such an extreme threshold.

First, it is not clear that the same significance threshold should be used in all cases. Whether one is convinced that a discovery is real should take into account the plausibility of the implied signal and how well it describes the data. If the discovered phenomenon is a priori very unlikely, then more evidence is required to produce a given degree of belief that the new phenomenon exists. As Carl Sagan said, "... extraordinary claims require extraordinary evidence" [44]. This follows directly from Bayes' theorem (36), whereby the posterior probability of a hypothesis is proportional to its prior probability. If an experimental result can only be explained by phenomena that may not be impossible but nevertheless highly improbable (fifth force, superluminal neutrinos), then it seems natural to demand a higher level of statistical significance.

Some phenomena, on the other hand, are regarded by the community as quite likely to exist before they are observed experimentally. Most particle physicists would have bet on the Higgs boson well in advance of the direct experimental evidence. As with the Higgs, however, when a discovery is announced in HEP it is usually something fairly important and the cost of an incorrect claim is perceived to be quite high. Every time the community endures a false discovery there is a tendency to think that the threshold should be higher.

Another reason for the high five-sigma threshold is that the experimenter may be unsure of the statistical model on which the reported significance relies. To first approximation one can think of the significance Z as the estimated size of the signal divided by the standard deviation σ in the estimated background. Here σ characterizes the level of random fluctuation in the background, i.e., it is a statistical error. If we have a systematic uncertainty in the background as well, then roughly speaking these should get added in quadrature. If an underestimate of our systematic errors would result in our σ being wrong by a factor of several, then a mere three-sigma effect may be no real effect at all. The high threshold in this case thus compensates for modeling uncertainty.

Another important issue is the look-elsewhere effect, where as discussed in Sect. 12 it is difficult to define exactly where else one looked. That is, should one correct for the fact that the search histogram had 100 bins, or also for the fact that one looked at 100 different histograms, or perhaps account for the thousands of scientists all carrying out searches? Surely in such a scenario someone will see a bump in a histogram somewhere that appears significant. Since it is impossible to draw an unambiguous boundary around where one "looked", there always remains a nagging feeling that one's correction for this effect may have been inadequate, hence the desire for a greater margin of safety before announcing a discovery.

The p-value, however, really only addresses the issue of whether a fluctuation in the background-only model is likely to lead to data as dissimilar to background as what was actually obtained. It is not designed to compensate for systematic errors in the model, the cost of announcing a false discovery or the plausibility of the phenomena implied by the discovery. Usually when a new phenomenon is discovered, it appears initially as only marginally significant, then continues to emerge until everyone is convinced. At first, everyone asks whether the apparent signal is just a fluctuation, but at some point people stop asking that question, because it is obvious that something has been observed. The question is whether that something is "new physics" or an uncontrolled systematic effect. Provided that the look-elsewhere effect is taken into account in a reasonable way, this transition probably takes place closer to the three-sigma level, in any case well before $Z = 5$.

Nevertheless, the 5-sigma threshold continues to be used to decide when the word "discovery" is appropriate. In future the HEP community should perhaps think of better ways of answering the different questions that arise when searching for new phenomena, since the statistical significance is really only designed to say whether the data, in the absence of a signal, is likely to have fluctuated in manner at least as extreme as what was observed. Lumping all of the issues mentioned above into the p-value simply makes them more difficult to disentangle.

15 Conclusions

To discover a new physical phenomenon we need to be able to demonstrate quantitatively that our data cannot be described using only known processes. In these lectures we have seen how statistical tests allow us to carry out this task. They provide a framework for rejecting hypotheses on the basis that the data we observed were uncharacteristic for them and more indicative of an alternative explanation. Frequentist statistical tests nevertheless prevent one from asking directly certain seemingly relevant questions, such as "what is the probability that my theory is true?". Bayesian statistics does allow one to quantify such a degree of belief, at the expense of having to supply subjective prior probabilities. The frequentist and Bayesian approaches answer different, but related questions, and both are valuable tools.

We did not have time to discuss in detail many other statistical issues such as Bayesian methods for establishing discovery, multivariate techniques and more sophisticated means for improving the accuracy of statistical models by introducing carefully motivated nuisance parameters. These methods will no doubt play an important role when the LHC enters its next data-taking phase.

Acknowledgements I wish to convey my thanks to the students and organizers of the 69th SUSSP in St. Andrews for the stimulating environment, friendly atmosphere and lively discussions that resulted in a highly productive and enjoyable school.

References

1. J. Beringer et al., (Particle Data Group), The review of particle physics. Phys. Rev. **D86**, 010001 (2012)
2. G.D. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998)
3. L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, Cambridge, 1986)
4. R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley, Chichester, 1989)
5. F. James, *Statistical Methods in Experimental Physics*, 2nd edn. (World Scientific, Singapore, 2006)
6. S. Brandt, *Data Analysis*, 3rd edn. (Springer, New York 1999)
7. G. Cowan, A survey of unfolding methods for particle physics, in *Proceedings of Conference on Advanced Statistical Techniques in Particle Physics*, IPPP/02/39, Durham, 2002, ed. by M.R. Whalley, L. Lyons
8. A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933); *Foundations of the Theory of Probability*, 2nd edn. (Chelsea, New York, 1956)
9. A. Stuart, J.K. Ord, S. Arnold, *Kendall's Advanced Theory of Statistics, Vol. 2A: Classical Inference and the Linear Model*, 6th edn. (Oxford University Press, Oxford, 1999), and earlier editions by Kendall and Stuart
10. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006)
11. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, 2nd edn. (Springer, New York, 2009)
12. R. Duda, P. Hart, D. Stork, *Pattern Classification*, 2nd edn. (Wiley, New York, 2001)
13. A. Webb, *Statistical Pattern Recognition*, 2nd edn. (Wiley, Chichester, 2002)
14. G. Cowan, Topics in statistical data analysis for HEP, in *LHC Physics*, ed. by T. Binoth, C. Buttar, P.J. Clark, E.W.N. Glover (Taylor and Francis, Boca Raton, 2012) (see also arXiv:1012.3589)
15. Links to the Proceedings of the PHYSTAT conference series (Durham 2002, Stanford 2003, Oxford 2005, and Geneva 2007) can be found at `phystat.org`
16. I. Narsky, StatPatternRecognition: a C++ package for statistical analysis of high energy physics data, `arXiv:physics/0507143`; software available from `statpatrec.sourceforge.net`
17. A. Höcker et al., TMVA users guide, `arXiv:physics/0703039`; software available from `tmva.sourceforge.net`
18. H. Hu, J. Nielsen, Analytic confidence level calculations using the likelihood ratio and fourier transform, arXiv:physics/9906010 [physics.data-an] (1999)
19. T.A. Severini, Likelihood ratio statistics based on an integrated likelihood. Biometrika **97**, 481 (2010)
20. R.E. Kass, A.E. Raftery, Bayes factors. J. Am. Stat. Assoc. **90**(430), 773 (1995)
21. T. Aaltonen et al., Combination of Tevatron searches for the standard model Higgs boson in the W^+W^- decay mode. Phys. Rev. Lett. **104**, 061802 (2010)
22. S.S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. **9**, 60 (1938)
23. A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Am. Math. Soc. **54**, 426 (1943)
24. G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics. Eur. Phys. J. **C71**, 1554 (2011)
25. K. Cranmer, Statistical challenges for searches for new physics at the LHC, in *Proceedings of PHYSTAT 2005*, Oxford, 2005 ed. by L. Lyons, M. Karagoz Unel
26. C. Chuang, T.L. Lai, Stat. Sin. **10**, 1 (2000)
27. M. Walker B. Sen, M. Woodroofe, Stat. Sin. **19**, 301 (2009)

28. R.D. Cousins, G.J. Feldman, A unified approach to the classical statistical analysis of small signals. Phys. Rev. **D57**, 3873 (1998)
29. R.E. Kass, L. Wasserman, The selection of prior distributions by formal rules. J. Am. Stat. Assoc. **91**(435), 1343 (1996)
30. L. Demortier, S. Jain, H.B. Prosper, Reference priors for high energy physics. Phys. Rev. **D82**, 034002 (2010)
31. D. Casadei, Reference analysis of the signal + background model in counting experiments. JINST **7**, P01012 (2012)
32. C.P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd edn. (Springer, New York, 2004)
33. J.S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001)
34. R.M. Neal, Probabilistic inference using Markov Chain Monte Carlo methods, Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, available from www.cs.toronto.edu/~radford/res-mcmc.html
35. K. Cranmer, Frequentist hypothesis testing with background uncertainty, in *Proceedings of PHYSTAT 2003*, ed. by L. Lyons et al. SLAC, Stanford, 8–11 Sept 2003, p. 261
36. L. Demortier, Objective Bayesian upper limits for Poisson processes, CDF Memo 5928, 2005
37. A.L. Read, Presentation of search results: the CL_s technique. J. Phys. G **28**, 2693 (2002)
38. T. Junk, Confidence level computation for combining searches with small statistics. Nucl. Instrum. Methods Phys. Res. Sect. A **434**, 435 (1999)
39. E. Gross, O. Vitells, Trial factors for the look elsewhere effect in high energy physics. Eur. Phys. J. **C70**, 525 (2010)
40. E. Gross, O. Vitells, Estimating the significance of a signal in a multi-dimensional search. Astropart. Phys. **35**, 230 (2011)
41. R.B. Davis, Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74**, 33 (1987)
42. G. Aad et al., ATLAS Collaboration, 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 (2012)
43. S. Chatrchyan et al., CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. **B 716**, 30 (2012)
44. C. Sagan, *Cosmos*, Episode 12, PBS (14 Dec 1980)