

**Particle Physics**  
*(Course of Lectures)*

**Merab GOGBERASHVILI**

**Javakhishvili Tbilisi State University**  
Faculty of Exact and Natural Sciences  
*Department of Physics*

**Tbilisi, 2018**



# Contents

<b>I Lecture – Review of Classical Physics</b>	<b>1</b>
<b>1 Some Concepts</b>	<b>3</b>
1.1 Hamilton’s Principle . . . . .	3
1.2 Noether’s Theorem . . . . .	4
1.2.1 Conservation of Momentum . . . . .	5
1.2.2 Conservation of Energy . . . . .	6
<b>2 Relativistic Field Equations</b>	<b>7</b>
2.1 Maxwell Equations . . . . .	7
2.1.1 Gradient Invariance . . . . .	9
2.2 Klein-Gordon Equation . . . . .	10
2.3 Proca Equation . . . . .	10
2.4 Weyl Equation . . . . .	11
2.4.1 Majorana Mass . . . . .	13
2.5 Dirac Equation . . . . .	14
2.6 Rarita-Schwinger Equation . . . . .	16

<b>3 Forces and Particles</b>	<b>19</b>
3.1 The Fundamental Forces . . . . .	19
3.2 Fundamental Particles . . . . .	20
3.2.1 Elementary Fermions . . . . .	21
3.2.2 Elementary Bosons . . . . .	24
3.3 Periodic Table of Particle Physics . . . . .	25
3.4 Composite Particles . . . . .	26
<b>II Lecture – Particle Reactions</b>	<b>29</b>
<b>4 Elements of the Scattering Theory</b>	<b>31</b>
4.1 Kinematics of the Minkowski Spacetime . . . . .	31
4.2 Collider vs Fixed Target Experiment . . . . .	34
4.2.1 Mandelstam Variables . . . . .	34
4.2.2 Symmetric Collisions ( $A + A$ ) . . . . .	35
4.2.3 Asymmetric Collisions ( $A + B$ ) . . . . .	37
<b>5 Golden Rules for Particle Reactions</b>	<b>39</b>
5.1 Free Field Solutions . . . . .	39
5.1.1 Spin 0: Scalar Field . . . . .	39
5.1.2 Spin 1/2: the Dirac Field . . . . .	40
5.1.3 Spin 1: Vector Field . . . . .	42
5.2 Differential Cross Section . . . . .	43
5.3 Fermi Golden Rules . . . . .	45



5.4	Feynman Diagrams for $\phi^4$ -model . . . . .	47
5.5	Calculating $e^+e^- \rightarrow \mu^+\mu^-$ in QED . . . . .	50
5.5.1	Summing Over Polarizations . . . . .	51
5.5.2	Casimir Trick . . . . .	52
5.5.3	Traces and Contraction Identities of $\gamma$ -Matrices . . . . .	52
5.5.4	Kinematics in the CM Frame . . . . .	53
5.5.5	Integration Over 2-Particle Phase Space . . . . .	54
5.5.6	Summary of Steps . . . . .	55
<b>6</b>	<b>Conservation Laws and Selection Rules</b> . . . . .	<b>57</b>
6.1	Crossing Symmetry . . . . .	57
6.2	Totalitarian Principle . . . . .	58
6.3	Conservation of Baryon Number . . . . .	60
6.4	Conservation of Lepton Number . . . . .	61
6.5	Isospin . . . . .	62
6.5.1	Strangeness . . . . .	63
6.5.2	Charm . . . . .	65
6.5.3	Beauty . . . . .	65
6.5.4	Truth . . . . .	65
6.6	$C$ , $P$ and $T$ Transformations . . . . .	66
6.6.1	Charge Conjugation . . . . .	66
6.6.2	Parity . . . . .	67
6.6.3	Time Reversal . . . . .	70
6.6.4	CP Invariance . . . . .	70

6.6.5	<i>CPT</i> Theorem . . . . .	71
<b>III</b>	<b>Lecture – Quantum Field Theory</b>	<b>73</b>
<b>7</b>	<b>Quantum Fields</b>	<b>75</b>
7.1	Primary and Secondary Quantizations . . . . .	75
7.2	Quantization of Spin-0 Fields . . . . .	77
7.3	Quantization of Spin 1/2 Fields . . . . .	78
7.3.1	The Dirac Sea Interpretation of Antiparticles . . . . .	81
7.4	The QFT Interpretation of Antiparticles . . . . .	82
7.5	Dirac Equation with Electromagnetic Field . . . . .	83
<b>8</b>	<b>Canonical Quantization</b>	<b>85</b>
8.1	Canonical Quantization of Scalar Fields . . . . .	86
8.2	The Spin-Statistics Theorem . . . . .	91
8.2.1	Left-Handed and Right-Handed Fields . . . . .	92
8.3	Canonical Quantization of Fermions . . . . .	94
<b>9</b>	<b>Path Integral Quantization</b>	<b>97</b>
9.1	Interpretation of the Path Integral . . . . .	100
9.2	Expectation Values . . . . .	101
9.3	Path Integrals with Fields . . . . .	102
9.4	Interacting Scalars and Feynman Diagrams . . . . .	106
9.5	Interacting Fermion Fields . . . . .	111

<b>IV Lecture – Regularizations and Renormalizations</b>	<b>113</b>
<b>10 Regularization</b>	<b>117</b>
10.1 Realistic Regularization . . . . .	118
10.1.1 Historical Remarks and Opinions . . . . .	119
10.2 Dimensional Regularization . . . . .	121
10.3 Pauli-Villars Regularization . . . . .	123
10.4 Lattice Regularization . . . . .	123
10.5 Zeta Function Regularization . . . . .	124
10.5.1 Relation to Dirichlet Series . . . . .	127
10.5.2 Heat Kernel Regularization . . . . .	128
10.6 Causal Perturbation Theory . . . . .	128
10.7 Hadamard Regularization . . . . .	129
<b>11 Renormalization</b>	<b>131</b>
11.1 Conventional Renormalization . . . . .	132
11.1.1 Mass and Wavefunction Renormalization . . . . .	134
11.1.2 Coupling Constant Renormalization . . . . .	137
11.2 BPH Renormalization . . . . .	140
11.2.1 Power Counting Method . . . . .	141
11.2.2 More on BPH Renormalization . . . . .	143
11.3 Power Counting and Renormalizability . . . . .	146
11.3.1 Theories with Fermions and Scalar Fields . . . . .	147
11.3.2 Theories with Vector Fields . . . . .	151

11.3.3 Renormalization of Composite Operators . . . . .	152
11.3.4 Symmetry and Renormalization . . . . .	155
11.3.5 Ward Identity . . . . .	156
<b>12 Renormalization Group</b>	<b>159</b>
12.1 History of RG . . . . .	160
12.2 Block Spin RG . . . . .	163
12.3 Elementary Considerations . . . . .	165
12.4 Relevant Operators and Universality Classes . . . . .	166
12.5 Exact RG Equations . . . . .	167
<b>V Lecture – Introduction to Group Theory</b>	<b>171</b>
<b>13 Mathematical Descriptions of Groups</b>	<b>173</b>
13.1 Finite Discrete Groups . . . . .	175
13.2 Group Actions . . . . .	177
13.3 Representations . . . . .	179
13.4 Reducibility and Irreducibility . . . . .	182
13.5 Algebraic Definitions . . . . .	184
13.6 Reducibility Revisited . . . . .	189
<b>14 Lie Groups</b>	<b>195</b>
14.1 Classification of Lie Groups . . . . .	196
14.2 Generators . . . . .	201
14.3 Lie Algebras . . . . .	202

14.4 The Adjoint Representation . . . . .	205
14.5 Root Space . . . . .	206
14.6 Casimir Operators . . . . .	212
<b>15 Examples of Lie Groups</b>	<b>215</b>
15.1 $SO(2)$ . . . . .	216
15.2 $SO(3)$ . . . . .	217
15.3 $SU(2)$ . . . . .	218
15.3.1 $SU(2)$ and Physical States . . . . .	219
15.3.2 $SU(2)$ for $j = 1/2$ . . . . .	223
15.3.3 $SU(2)$ for $j = 1$ . . . . .	225
15.3.4 $SU(2)$ for Arbitrary $j$ . . . . .	225
15.3.5 Adjoint Representation of $SU(2)$ . . . . .	227
15.3.6 $SU(2)$ for Arbitrary $j$ ... Again . . . . .	230
15.4 $SU(3)$ . . . . .	231
<b>VI Lecture – Particle Experiments</b>	<b>235</b>
<b>16 Anatomy of Experiments</b>	<b>237</b>
16.1 Accelerators and Laboratories . . . . .	237
16.1.1 Luminosity . . . . .	238
16.1.2 Particle Physics Laboratories . . . . .	239
16.1.3 What is After LHC? . . . . .	240
16.2 Detectors . . . . .	241

16.2.1	Tracking Detectors . . . . .	243
16.2.2	Electromagnetic Calorimeters . . . . .	245
16.2.3	Hadron Calorimeters . . . . .	246
16.2.4	Muon Detectors . . . . .	247
16.2.5	Particle Identification . . . . .	249
16.2.6	Selection of Events and Triggers . . . . .	251
16.3	Data Analysis . . . . .	253
16.3.1	Monte Carlo . . . . .	253
16.3.2	What Can Go Wrong . . . . .	256
<b>17</b>	<b>Limitations of Particle Accelerators</b>	<b>259</b>
17.1	Limit I: The Geography . . . . .	259
17.2	Limit II: Voltage Breakdown in DC Accelerators . . . . .	260
17.3	Limit III: Size of the AC Accelerating Structures . . . . .	261
17.4	Limit IV: Magnetic Guide Field . . . . .	264
17.5	Limit V: Energy of Fixed-Target Colliders . . . . .	275
17.6	Limit VI: The Unavoidable Particle Detectors . . . . .	277
17.7	Limit VII: The Rareness of Searched Events . . . . .	279
17.8	Limit VIII: Luminosity of a Collider Ring . . . . .	280
17.9	Limit IX: Power Loss in Synchrotron Radiation . . . . .	282
17.10	Limit X: Gradients in Linear Structures . . . . .	284
<b>18</b>	<b>History of Validation of SM</b>	<b>287</b>
18.1	Discovery of Neutral Currents . . . . .	287

18.2	The Discovery of Neutrinoless NC . . . . .	295
18.3	Observation of Charm . . . . .	299
18.4	Fixing of $\sin^2 \theta_W$ and Prediction of $M_W, M_Z$ . . . . .	300
18.5	Direct Evidence for the $ZWW$ Coupling . . . . .	307
18.6	Precision Testing of SM . . . . .	308
18.6.1	Radiative Corrections and $\rho/\sin^2 \theta_W$ . . . . .	309
18.6.2	Precision Measurements at LEP . . . . .	312
18.6.3	Precision Testing and Indirect Bounds . . . . .	315
18.7	Mass of the Higgs . . . . .	322
18.7.1	Unitarity Bound . . . . .	322
18.7.2	Triviality Bound . . . . .	324
18.7.3	Stability Bound . . . . .	326
18.8	Discovery of the Higgs Boson . . . . .	332
<b>VII</b>	<b>Lecture – Introduction to the Standard Model</b>	<b>337</b>
<b>19</b>	<b>Gauge Interactions</b>	<b>339</b>
19.1	Gauging the Symmetry . . . . .	339
19.2	Non-Abelian Case . . . . .	343
19.3	Representations of Gauge Groups . . . . .	345
<b>20</b>	<b>The Unification Scheme</b>	<b>347</b>
20.1	Fermi’s Theory for Weak Interactions . . . . .	347
20.2	Decay Constants . . . . .	348

20.3 Quark Mixing . . . . .	349
20.4 Intermediate Bosons . . . . .	351
<b>21 Particle Content and Currents</b>	<b>355</b>
21.1 Gauge Bosons . . . . .	355
21.2 Fermions . . . . .	356
21.3 Neutral Bosons Mixing . . . . .	357
21.4 Currents . . . . .	358
21.5 Strengths of the CC and NCs . . . . .	361
<b>VIII Lecture – Spontaneous Symmetry Breaking</b>	<b>365</b>
<b>22 The Mechanism of SSB</b>	<b>367</b>
22.1 The Beginnings of SSB . . . . .	367
22.2 Breaking Global $U(1)$ . . . . .	370
22.3 Breaking Local $U(1)$ . . . . .	372
22.4 SSB in Non-Abelian Gauge Theory . . . . .	373
22.5 Simple Examples of SSB . . . . .	375
22.6 Breaking of $SU(N)$ . . . . .	377
<b>23 SSB in Condensed Matter</b>	<b>379</b>
23.1 Ginzburg-Landau Model of Superconductivity . . . . .	379
23.2 High- $T_c$ Superconductors . . . . .	380
23.3 Superfluid $^3He$ . . . . .	381
23.4 Superfluid $^4He$ : . . . . .	382



<i>CONTENTS</i>	xiii
23.5 Other Systems . . . . .	383
23.6 SSB in Classical Mechanics . . . . .	383
23.6.1 Field Equations . . . . .	386
23.6.2 SSB and Higgs Mechanism . . . . .	387
<b>24 Higgs Mechanism in SM</b>	<b>389</b>
24.1 Gauge Bosons Masses . . . . .	389
24.2 The Higgs Sector in Lagrangian . . . . .	394
24.3 Yukawa Interactions . . . . .	395
<b>IX Lecture – Summary of the SM and Predictions</b>	<b>399</b>
<b>25 The SM Lagrangian</b>	<b>401</b>
25.1 The Gauge Part . . . . .	401
25.2 The Scalar Part . . . . .	402
25.2.1 Masses of Gauge Bosons . . . . .	404
25.3 The Fermion Part . . . . .	407
25.3.1 The Lepton Sector . . . . .	407
25.3.2 The Quark Sector . . . . .	410
25.4 The Yukawa Couplings . . . . .	412
<b>26 Interactions</b>	<b>415</b>
26.1 Charged Currents . . . . .	415
26.1.1 QED Part . . . . .	417
26.2 Neutral Currents . . . . .	418

26.3 Anomaly Cancellation . . . . .	420
<b>27 Predictions of the SM</b>	<b>423</b>
27.1 High Energy Scattering . . . . .	424
27.2 Observations Meet Predictions . . . . .	427
27.3 Free Parameters in the SM . . . . .	429
<b>X Lecture – Introduction to QCD</b>	<b>433</b>
<b>28 History and Outlook</b>	<b>435</b>
28.1 The Yukawa Model . . . . .	436
28.2 The Sakata Model . . . . .	438
28.3 The Eightfold Way . . . . .	439
28.3.1 The Fundamental Representation: Quarks . . . . .	441
28.4 The OZI Rule . . . . .	442
28.5 Quark Approaches . . . . .	444
28.5.1 Heavy Quarks . . . . .	444
28.5.2 A First Hint of Colour . . . . .	446
28.6 Development of QCD . . . . .	447
28.7 Distinguished Properties . . . . .	448
28.7.1 Color Confinement . . . . .	448
28.7.2 Asymptotic Freedom and Duality . . . . .	449
28.8 The Strong Coupling . . . . .	450
28.9 Methods of QCD . . . . .	452

<i>CONTENTS</i>	xv
28.10 Confirmation and Outlook . . . . .	454
<b>29 Structure of QCD</b>	<b>457</b>
29.1 Symmetries of the Model . . . . .	457
29.1.1 Color States . . . . .	459
29.2 The Lagrangian of QCD . . . . .	461
29.3 Color Factors . . . . .	464
29.4 Renormalization of QCD . . . . .	465
29.4.1 The Example from QED . . . . .	466
29.4.2 RG in QCD . . . . .	467
<b>30 Symmetry Breaking Patterns in QCD</b>	<b>471</b>
30.1 Chiral Symmetry in QCD . . . . .	471
30.1.1 Chiral SSB . . . . .	474
30.2 Hadron Masses . . . . .	478
30.3 The Example of $\sigma$ -model . . . . .	480
30.3.1 Lagrangian of the $\sigma$ -model . . . . .	480
30.3.2 Pion-pion Scattering at Tree Level . . . . .	483
30.3.3 Large $-M_\sigma$ Limit in the Linear $\sigma$ -model . . . . .	485
<b>XI Lecture – Beyond the SM</b>	<b>489</b>
<b>31 Problems with SM</b>	<b>491</b>
31.1 The Gauge Problem . . . . .	492
31.2 The Fermion Problem . . . . .	493

31.3	The Hierarchy Problem . . . . .	494
31.4	The Strong $CP$ Problem . . . . .	497
31.5	The Gravity Problem . . . . .	498
31.6	The New Ingredients . . . . .	500
<b>32</b>	<b>Flavor and <math>CP</math> Violation in EFT</b>	<b>505</b>
32.1	The Flavor Sector of the SM . . . . .	506
32.1.1	Meson Mixing and the GIM Mechanism . . . . .	509
32.2	$CP$ Violation in Meson Decays . . . . .	510
32.3	Flavor Transitions . . . . .	514
32.3.1	The New Physics Flavor Puzzle . . . . .	516
32.3.2	The Minimal Flavor Violation Ansatz . . . . .	518
32.3.3	EFTs for Rare $B$ Decays . . . . .	520
32.3.4	Top and Higgs Flavor Violating Signatures . . . . .	522
<b>33</b>	<b>Neutrino Mass Problem</b>	<b>525</b>
33.1	Brief Historical Survey . . . . .	525
33.2	Neutrino Mixing . . . . .	527
33.3	Flavor Neutrino States . . . . .	530
33.4	Neutrino Oscillations in Vacuum . . . . .	531
33.5	Neutrino and SM . . . . .	534
33.6	The Weinberg Mechanism of Mass Generation . . . . .	535

**XII Lecture – Supersymmetry 539**

**34 Overview and Current Status 541**

34.1 History of SUSY Invention . . . . . 542

34.2 Motivations . . . . . 543

34.3 Extension of Symmetry Groups . . . . . 545

34.4 Applications . . . . . 546

34.4.1 The Supersymmetric SM . . . . . 546

34.4.2 Supersymmetric Quantum Mechanics . . . . . 547

34.4.3 SUSY in Condensed Matter Physics . . . . . 548

34.4.4 SUSY in Optics . . . . . 548

34.4.5 SUSY in Dynamical Systems . . . . . 549

34.4.6 SUSY in Mathematics . . . . . 549

34.4.7 SUSY in Quantum Gravity . . . . . 550

34.5 Extended SUSY . . . . . 550

34.5.1 SUSY and Extra Dimensions . . . . . 551

34.5.2 Fractional SUSY . . . . . 551

34.6 Experimental Searches . . . . . 551

34.6.1 SUSY Signatures on LHC . . . . . 552

**35 Main Concepts 555**

35.1 Spacetime Symmetry . . . . . 556

35.2 A Simple SUSY Model . . . . . 559

35.3 The Vacuum Energy . . . . . 561

35.4	SUSY Breaking . . . . .	569
35.4.1	The O’Raifeartaigh Model . . . . .	569
35.5	SUSY Lagrangians . . . . .	571
35.5.1	A SUSY Gauge Theory . . . . .	572
35.5.2	Yukawa-like Interactions . . . . .	573
35.5.3	$R$ -symmetry . . . . .	574
35.5.4	$F$ -terms and $D$ -terms . . . . .	574
35.5.5	Local SUSY and Supergravity . . . . .	575
<b>36</b>	<b>The SUSY and SM</b>	<b>577</b>
36.1	Motivation . . . . .	577
36.2	The Structure of the Model . . . . .	578
36.2.1	$R$ -symmetry . . . . .	583
36.3	SUSY Breaking . . . . .	585
36.3.1	Breaking Terms . . . . .	585
36.4	Mediating the Breaking . . . . .	586
36.4.1	Gauge Mediated Breaking . . . . .	586
36.4.2	Gravity Mediation . . . . .	589
36.4.3	Anomaly Mediation . . . . .	589
36.5	Planck Suppressed Operators . . . . .	590
36.6	EW SSB and the Higgs Mass . . . . .	591
36.6.1	Neutralinos and Charginos . . . . .	595
36.6.2	Sfermion Spectrum . . . . .	596
36.6.3	Flavor Structure . . . . .	597

36.6.4 <i>R</i> -parity Violating Couplings . . . . .	598
<b>XIII Lecture – Grand Unification Theories</b>	<b>599</b>
<b>37 Preliminaries</b>	<b>601</b>
37.1 GUT Idea . . . . .	602
37.1.1 Historical Remarks . . . . .	607
37.1.2 Gauge Coupling Unification . . . . .	610
37.2 The Pati-Salam Model . . . . .	611
37.2.1 Left-Right Asymmetry in SM . . . . .	612
37.2.2 Left-Right Symmetry . . . . .	613
37.2.3 Symmetry Breaking . . . . .	615
37.2.4 The Origin of Neutrino Masses . . . . .	617
37.2.5 Lepton Number as a 4-th Color . . . . .	619
37.2.6 Nuclear Stability . . . . .	619
<b>38 Georgi-Glashow’s <math>SU(5)</math></b>	<b>621</b>
38.1 Structure of the Model . . . . .	621
38.1.1 Charge Quantization . . . . .	623
38.1.2 Anomaly Cancellation . . . . .	623
38.1.3 Gauge Coupling Running . . . . .	624
38.2 Symmetry Breaking . . . . .	626
38.2.1 The First Stage . . . . .	626
38.2.2 The Second Stage and Doublet-Triplet Splitting . . . . .	629

38.3 Proton Decay . . . . .	630
38.4 Yukawa Sector and Neutrinos . . . . .	631
38.4.1 Neutrino Masses . . . . .	633
38.5 SUSY Extension . . . . .	634
<b>39 <math>SO(10)</math> GUT</b>	<b>637</b>
39.1 $SO(10)$ Group . . . . .	639
39.1.1 Tensor Representations . . . . .	640
39.1.2 Spinor Representations . . . . .	641
39.2 Anomaly Cancellation . . . . .	648
39.3 The SM Embedding . . . . .	648
39.4 The Higgs Sector . . . . .	650
<b>XIV Afterwards</b>	<b>653</b>



# Notations

## List of Abbreviations

**BSM** – Beyond the Standard Model

**CC** – Charged Current

**EFT** – Effective Field Theory

**eV** – Electron Volt

**EW** – Electro Weak

**GeV** – Giga Electron Volt

**GUT** – Grand Unification Theory

**LHC** – Large Hadron Collider

**MFV** – Minimal Flavor Violation

**MSSM** – Minimal Supersymmetric Standard Model

**NC** – Neutral Current

**QCD** – Quantum Chromodynamics

**QFT** – Quantum Field Theory

**SM** – Standard Model of Particle Physics

**SSB** – Spontaneous Symmetry Braking

**SUSY** – Supersymmetry

**UV** – Ultraviolet

**VEV** – Vacuum Expectation Value

**WIMP** – Weakly Interacting Massive Particles

**WKB** – Wentzel–Kramers–Brillouin

## Conventions

In most cases the system of natural units is used:  $\hbar = c = 1$ .

The 4-dimensional metric has the signature:  $(+, -, -, -)$ .

The Levi-Civita tensor:  $\varepsilon_{0123} = -\varepsilon^{0123} = 1$ .

Dirac's derivatives and momentums:  $\not{\partial} = \gamma^\nu \partial_\nu$ ,  $\not{p} = \gamma^\nu p_\nu$ .

Repeated lower and upper indices are summed over, unless otherwise stated.

## Part I

# Lecture – Review of Classical Physics



# Chapter 1

## Some Concepts

### 1.1 Hamilton's Principle

Nearly all physics begins with what is called a *Lagrangian* for a particle, which is initially defined as the kinetic energy minus the potential energy,

$$L \equiv T - V , \tag{1.1}$$

where  $T = T(q, \dot{q})$  and  $V = V(q)$ . Then, the *Action* is defined as the integral of the Lagrangian from an initial time to a final time,

$$S \equiv \int_{t_i}^{t_f} dt L(q, \dot{q}) . \tag{1.2}$$

It is important to realize that  $S$  is a 'functional' of the particle's world-line in  $(q, \dot{q})$  space, not a function. This means that it depends on the entire path  $(q, \dot{q})$ , rather than a given point on the path. The only fixed points on the path are  $q(t_i)$ ,  $q(t_f)$ ,  $\dot{q}(t_i)$ , and  $\dot{q}(t_f)$ . The rest of the path is generally unconstrained, and the value of  $S$  depends on the entire path.

*Hamilton's Principle* says that nature extremizes the path a particle will take in going from  $q(t_i)$  at time  $t_i$  to position  $q(t_f)$  at time  $t_f$ . In other words, the path that extremizes the action will be the path the particle will travel. But, because  $S$  is a functional, depending on the entire path in  $(q, \dot{q})$  space rather than a point, it cannot be extremized setting the derivative equal to

0. Instead, we must find the path for which the action is 'stationary'. This means that the first-order term in the Taylor Expansion around that path will vanish, or  $\delta S = 0$  at that path.

To find this, consider some arbitrary path  $(q, \dot{q})$ . If it is a path that minimizes the action, then we will have

$$\begin{aligned}\delta S &= \delta \int_{t_i}^{t_f} dt L(q, \dot{q}) = \int_{t_i}^{t_f} dt L(q + \delta q, \dot{q} + \delta \dot{q}) - S = \\ &= \int_{t_i}^{t_f} dt L(q, \dot{q}) + \int_{t_i}^{t_f} dt \left( \delta q \frac{\partial L}{\partial q} + \delta \dot{q} \frac{\partial L}{\partial \dot{q}} \right) - S = \\ &= \int_{t_i}^{t_f} dt \left( \delta q \frac{\partial L}{\partial q} + \frac{\partial L}{\partial \dot{q}} \frac{d}{dt} \delta q \right) = 0 .\end{aligned}\tag{1.3}$$

Integrating the second term by parts, and taking the variation of  $\delta q$  to be at 0 at  $t_i$  and  $t_f$ ,

$$\delta S = \int_{t_i}^{t_f} dt \left( \delta q \frac{\partial L}{\partial q} - \delta q \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) = \int_{t_i}^{t_f} dt \delta q \left( \frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) = 0 .\tag{1.4}$$

The only way to guarantee this for an arbitrary variation  $\delta q$  from the path  $(q, \dot{q})$  is to demand

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0 .\tag{1.5}$$

This equation is called the *Euler-Lagrange* equation, and it produces the equations of motion of the particle. The generalization to multiple coordinates  $q_i$  ( $i = 1, \dots, n$ ) is straightforward:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 .\tag{1.6}$$

## 1.2 Noether's Theorem

Given a Lagrangian  $L = L(q, \dot{q})$ , consider an infinitesimal transformation

$$q \rightarrow q + \epsilon \delta q ,\tag{1.7}$$

where  $\epsilon$  is some infinitesimal constant. This transformation will give

$$L(q, \dot{q}) \rightarrow L(q + \epsilon \delta q, \dot{q} + \epsilon \delta \dot{q}) = L(q, \dot{q}) + \epsilon \delta q \frac{\partial L}{\partial q} + \epsilon \delta \dot{q} \frac{\partial L}{\partial \dot{q}} .\tag{1.8}$$

If the Euler-Lagrange equations (1.6) are satisfied, then under  $q \rightarrow q + \epsilon \delta q$ ,

$$\begin{aligned} L &\rightarrow L + \epsilon \delta q \frac{\partial L}{\partial q} + \epsilon \delta \dot{q} \frac{\partial L}{\partial \dot{q}} = \\ &= L + \epsilon \delta q \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} + \epsilon \frac{\partial L}{\partial \dot{q}} \frac{d}{dt} \delta q = L + \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \epsilon \delta q \right) . \end{aligned} \quad (1.9)$$

So, under  $q \rightarrow q + \epsilon \delta q$ , we have

$$\delta L = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \epsilon \delta q \right) . \quad (1.10)$$

We define the *Noether Current*,  $j$ , as

$$j \equiv \frac{\partial L}{\partial \dot{q}} \delta q . \quad (1.11)$$

Now, if we can find some transformation  $\delta q$  that leaves the action invariant, or in other words such that  $\delta S = 0$ , then

$$\frac{dj}{dt} = 0 , \quad (1.12)$$

and therefore the current  $j$  is a constant in time. In other words,  $j$  is conserved. So, *Noether's Theorem* says that whenever there is a continuous symmetry in the action, there is a corresponding conserved quantity.

### 1.2.1 Conservation of Momentum

As a familiar example, consider a projectile, described by the Lagrangian

$$L = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2) - mgy . \quad (1.13)$$

This will be unchanged under the transformation  $x \rightarrow x + \epsilon$ , where  $\epsilon$  is any constant (here,  $\delta q = 1$  in the above notation), because

$$x \rightarrow x + \epsilon \Rightarrow \dot{x} \rightarrow \dot{x} . \quad (1.14)$$

So,

$$j = \frac{\partial L}{\partial \dot{q}} \delta q = m\dot{x} \quad (1.15)$$

is conserved. We recognize  $m\dot{x}$  as the momentum in the  $x$ -direction, which we expect to be conserved by conservation of momentum.

### 1.2.2 Conservation of Energy

Consider the quantity

$$\frac{dL}{dt} = \frac{d}{dt}L(q, \dot{q}) = \frac{\partial L}{\partial q} \frac{dq}{dt} + \frac{\partial L}{\partial \dot{q}} \frac{d\dot{q}}{dt} + \frac{\partial L}{\partial t} . \quad (1.16)$$

Because  $L$  does not depend explicitly on time,  $\partial L/\partial t = 0$ , and therefore

$$\frac{dL}{dt} = \frac{\partial L}{\partial q} \dot{q} + \frac{\partial L}{\partial \dot{q}} \ddot{q} = \left( \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) \dot{q} + \frac{\partial L}{\partial \dot{q}} \ddot{q} = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \dot{q} \right) , \quad (1.17)$$

where we have used the Euler-Lagrange equation to get the second equality. So, we have

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \dot{q} - L \right) = 0 . \quad (1.18)$$

For a non-relativistic system (1.1),  $V$  is a function of  $q$  only, and normally

$$T \propto \dot{q}^2 \quad \Rightarrow \quad \frac{\partial L}{\partial \dot{q}} \dot{q} = 2T . \quad (1.19)$$

So, the total energy of the system,

$$\frac{\partial L}{\partial \dot{q}} \dot{q} - L = 2T - (T - V) = T + V = E , \quad (1.20)$$

is conserved according to (1.18). We identify

$$T + V \equiv H \quad (1.21)$$

as the *Hamiltonian*, or total energy function, of the system.

Furthermore, we define

$$\frac{\partial L}{\partial \dot{q}} \equiv p \quad (1.22)$$

to be the momentum of the system. Then, the relationship between the Lagrangian and the Hamiltonian is the Legendre transformation

$$p\dot{q} - L = H \quad (1.23)$$

**Exercise 1.1:** Consider a non-relativistic system of two particles interacting through a potential, which depends only on the relative coordinate  $r_1 - r_2$ . Show that the total momenta  $p = m_1 v_1 + m_2 v_2$  are conserved.

**Exercise 1.2:** Derive the Euler-Lagrange equation of motion for the case when Lagrangian also depends on the second derivatives of a field.



## Chapter 2

# Relativistic Field Equations

By relativistically invariant field equations we mean linear differential equations whose coefficients do not depend on  $x^\nu$  and invariant under the local Lorentz transformations. Some examples of relativistically invariant field equations used in physical applications are described below.

### 2.1 Maxwell Equations

We choose our units so that  $c = \mu_0 = \epsilon_0 = 1$ . So, the magnitude of the classical force between two electric charges  $q_1$  and  $q_2$  is

$$F = \frac{q_1 q_2}{4\pi r^2} . \tag{2.1}$$

In these units, classical Maxwell's equations are

$$\begin{aligned} \bar{\nabla} \cdot \bar{E} &= \rho , & \bar{\nabla} \times \bar{B} - \frac{\partial \bar{E}}{\partial t} &= \bar{J} , \\ \bar{\nabla} \cdot \bar{B} &= 0 , & \bar{\nabla} \times \bar{E} + \frac{\partial \bar{B}}{\partial t} &= 0 . \end{aligned} \tag{2.2}$$

If we define the *Potential* 4-vector  $A^\mu = (\phi, \bar{A})$ , then we can identify

$$\bar{B} = \bar{\nabla} \times \bar{A} , \quad \bar{E} = -\bar{\nabla} \phi - \frac{\partial \bar{A}}{\partial t} . \tag{2.3}$$

Writing  $\bar{B}$  and  $\bar{E}$  this way will automatically solve the homogenous Maxwell equations from the system (2.2).

Then, we define the totally antisymmetric *Electromagnetic Field Strength Tensor*  $F^{\mu\nu}$  as

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}, \quad (2.4)$$

and the 4-vector current as  $J^\mu = (\rho, \bar{J})$ . It is straightforward to show that

$$\begin{aligned} \partial^\lambda F^{\mu\nu} + \partial^\nu F^{\lambda\mu} + \partial^\mu F^{\nu\lambda} &= 0 \Rightarrow \left( \bar{\nabla} \cdot \bar{B} = 0 \ \& \ \bar{\nabla} \times \bar{E} + \frac{\partial \bar{B}}{\partial t} = 0 \right), \\ \partial_\mu F^{\mu\nu} &= J^\nu \Rightarrow \left( \bar{\nabla} \cdot \bar{E} = \rho \ \& \ \bar{\nabla} \times \bar{B} - \frac{\partial \bar{E}}{\partial t} = \bar{J} \right). \end{aligned} \quad (2.5)$$

Now let us construct a Lagrangian density  $\mathcal{L}$  which will, via Hamilton's Principle, produce Maxwell's equations. First, we know that  $\mathcal{L}$  must be a scalar (no uncontracted indices). The natural choice is to take  $A^\mu$  as the fundamental field and from classical physics we know that kinetic terms are quadratic in the derivatives of the fundamental coordinates. Therefore, the correct choice is

$$\mathcal{L}_{EM} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu. \quad (2.6)$$

So,

$$S = \int d^4x \left[ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu \right]. \quad (2.7)$$

Variation of (2.7) with respect to  $A^\mu$  gives,

$$\begin{aligned} \delta S &= \int d^4x \left[ -\frac{1}{4} F_{\mu\nu} \delta F^{\mu\nu} - \frac{1}{4} \delta F_{\mu\nu} F^{\mu\nu} - J^\mu \delta A_\mu \right] = \\ &= \int d^4x \left[ -\frac{1}{2} F_{\mu\nu} \delta F^{\mu\nu} - J^\mu \delta A_\mu \right] = \\ &= \int d^4x \left[ -\frac{1}{2} F_{\mu\nu} (\partial^\mu \delta A^\nu - \partial^\nu \delta A^\mu) - J^\mu \delta A_\mu \right] = \\ &= \int d^4x [-F_{\mu\nu} \partial^\mu \delta A^\nu - J^\mu \delta A_\mu]. \end{aligned} \quad (2.8)$$

Integrating the first term by parts, and choosing boundary conditions so that  $\delta A$  vanishes at the boundaries, we find:

$$\int d^4x [\partial_\mu F^{\mu\nu} \delta A_\nu - J^\nu \delta A_\nu] = \int d^4x [\partial_\mu F^{\mu\nu} - J^\nu] \delta A_\nu . \quad (2.9)$$

So, to have  $\delta S = 0$ , we must have  $\partial_\mu F^{\mu\nu} = J^\nu$ , and if this is written out one component at a time, it will give exactly the inhomogeneous Maxwell equations from the system (2.2). And as we already pointed out, the homogenous Maxwell equations become identities when written in terms of  $A^\mu$ .

### 2.1.1 Gradient Invariance

Given some specific potential  $A^\mu$ , we can find the field strength action as in (2.7). However,  $A^\mu$  does not uniquely specify the action. We can take any arbitrary function  $\chi(x^\mu)$ , and the action will be invariant under the gradient transformation

$$A^\mu \rightarrow A'^\mu = A^\mu + \partial^\mu \chi , \quad (2.10)$$

or

$$A^\mu \rightarrow A'^\mu = \left( \phi - \frac{\partial \chi}{\partial t}, \bar{A} + \bar{\nabla} \chi \right) . \quad (2.11)$$

Under this transformation, we have

$$\begin{aligned} F'^{\mu\nu} &= \partial^\mu A'^\nu - \partial^\nu A'^\mu = \partial^\mu (A^\nu + \partial^\nu \chi) - \partial^\nu (A^\mu + \partial^\mu \chi) = \\ &= \partial^\mu A^\nu - \partial^\nu A^\mu + \partial^\mu \partial^\nu \chi - \partial^\nu \partial^\mu \chi = F^{\mu\nu} . \end{aligned} \quad (2.12)$$

So,  $F'^{\mu\nu} = F^{\mu\nu}$ .

Furthermore,

$$J^\mu A_\mu \rightarrow J^\mu A'_\mu + J^\mu \partial_\mu \chi . \quad (2.13)$$

Integrating the second term by parts with the usual boundary conditions,

$$\int d^4x J^\mu \partial_\mu \chi = - \int d^4x (\partial_\mu J^\mu) \chi$$

But, according to Maxwell's equations,  $\partial_\mu J^\mu = \partial_\mu \partial_\nu F^{\mu\nu} \equiv 0$  because  $F^{\mu\nu}$  is totally antisymmetric. So, both  $F^{\mu\nu}$  and  $J^\mu \partial_\mu \chi$  are invariant under (2.10), and therefore the action of  $S$  is invariant under (2.10).

## 2.2 Klein-Gordon Equation

Let us consider a scalar field,  $\phi'(x') = \phi(x)$ . Using the notation

$$p_\mu = -i\hbar \frac{\partial}{\partial x^\mu}, \quad p^\mu = \eta^{\mu\nu} p_\nu, \quad (2.14)$$

the invariant Klein-Gordon equation can be written in the form:

$$-p_\mu p^\mu \phi(x) = m^2 c^2 \phi(x), \quad (2.15)$$

i.e.

$$\partial_\mu \partial^\mu \phi(x) = \frac{m^2 c^2}{\hbar^2} \phi(x). \quad (2.16)$$

The Klein-Gordon equation can be derived from the invariant action:

$$S = \int d^4x \left( \eta^{\mu\nu} \partial_\mu \phi^* \partial_\nu \phi + \frac{m^2 c^2}{\hbar^2} \phi^* \phi \right), \quad (2.17)$$

where  $\phi^*$  denotes complex conjugation of the field  $\phi$ .

Below, if otherwise is not stated, we shall use the universal system of units where  $c = \hbar = 1$ .

## 2.3 Proca Equation

The Proca equations,

$$\partial_\mu (\partial^\mu B^\nu - \partial^\nu B^\mu) - m^2 B^\nu = 0, \quad (2.18)$$

describes a complex massive spin-1 field  $B^\nu$ . This equation is equivalent to the conjunction of

$$(\partial_\mu \partial^\mu - m^2) B^\nu = 0, \quad (2.19)$$

with (in the massive case)

$$\partial_\mu B^\mu = 0, \quad (2.20)$$

which may be called a generalized Lorenz gauge condition.

The Proca equation is closely related to the Klein-Gordon equation, because it is second order in space and time. It is involved in the SM and describes

there the three massive vector bosons,  $Z^0$  and  $W^\pm$ . When  $m = 0$ , the equations (2.18) reduce to Maxwell's equations without charge or current.

The Proca (and Maxwell, when  $m = 0$ ) equations can be derived from the invariant action:

$$S = \int d^4x \left[ \frac{1}{2} (\partial_\mu B_\nu^* - \partial_\nu B_\mu^*) (\partial^\mu B^\nu - \partial^\nu B^\mu) - m^2 B_\nu^* B^\nu \right]. \quad (2.21)$$

## 2.4 Weyl Equation

We look for an equation whose field is represented by a two component vector which transforms accordingly to the  $(0, 1/2)$  representation

$$\psi'(x') = \tilde{A}\psi(x). \quad (2.22)$$

Here  $A\sigma_\mu x^\mu A^+ = \sigma_\mu x'^\mu$ , which is equivalent to  $\tilde{A}\tilde{\sigma}_\mu x^\mu \tilde{A}^+ = \tilde{\sigma}_\mu x'^\mu$ , where  $\tilde{A} = (A^+)^{-1}$  and  $\tilde{\sigma}_\mu = (\sigma_0, -\sigma_i)$ , with  $\sigma_0$  the identity and  $\sigma_i$  the Pauli matrices.

Such an equation is given by

$$\sigma_\mu p^\mu \psi(x) = 0. \quad (2.23)$$

In fact if (2.23) holds we have

$$\sigma_\mu p'^\mu \tilde{A}\psi(x) = \sigma_\mu p'^\mu \psi'(x') = 0, \quad (2.24)$$

as seen by rewriting (2.23) as

$$0 = A\sigma_\mu p^\mu A^+ \tilde{A}\psi = \sigma_\mu p'^\mu \tilde{A}\psi. \quad (2.25)$$

Multiplying on the left by  $\tilde{\sigma}_\mu p^\mu$  we have

$$-p_\mu p^\mu \psi(x) = 0. \quad (2.26)$$

So the Weyl equation describes particles of zero mass, since it is not possible to add to it a mass term,

$$\sigma_\mu p^\mu \psi(x) = M\psi(x), \quad (2.27)$$

$M$  being a  $2 \times 2$  matrix, without violating Lorentz invariance. In fact if  $\psi(x) \in (0, 1/2)$  we have  $\sigma_\mu p^\mu \psi(x) \in (1/2, 0)$  because

$$\sigma_\mu p'^\mu \psi'(x') = A\sigma_\mu p^\mu A^+ \tilde{\psi}(x) = A\sigma_\mu p^\mu \psi(x), \quad (2.28)$$

from which

$$M\psi(x) = A^{-1}M\tilde{A}\psi(x); \quad \forall A \in SL(2, C), \quad (2.29)$$

or

$$M\psi(x) = A^+MA\psi(x); \quad \forall A \in SL(2, C). \quad (2.30)$$

Choose a point  $x$  such that  $\psi(x) \neq 0$  and replacing  $A$  with  $AU$ ,  $A \in SL(2, C)$ ,  $U \in SU(2)$  we have

$$M\psi(x) = U^+A^+MAU\psi(x). \quad (2.31)$$

Taking the scalar product

$$\begin{aligned} (U\psi(x), A^+MAU\psi(x)) &= (\psi(x), M\psi(x)) \equiv \\ &\equiv c(\psi(x), \psi(x)) = c(U\psi(x), U\psi(x)) \end{aligned} \quad (2.32)$$

for all  $U \in SU(2)$  and as  $U\psi(x)$  covers the whole two dimensional space we have  $A^+MA = cI$ , which has to hold for all  $A \in SL(2, C)$ . For  $A = I$  we have  $M = cI$  so that  $cI = cA^+A$  for all  $A \in SL(2, C)$ , which implies it implies  $c = 0$ .

Going over to the Fourier transform we have

$$\sigma_\mu p^\mu \psi_f(p^0, \mathbf{p}) = 0, \quad \text{with} \quad (p^0)^2 - \mathbf{p}^2 = 0. \quad (2.33)$$

Equation (2.23) is not invariant under parity. In fact let us look for an invertible matrix  $P$  such that

$$\psi'(x') = P\psi(x) \quad \text{with} \quad x' = (x^0, -\mathbf{x}) \quad (2.34)$$

and such that

$$\sigma_\mu p'^\mu \psi'(x') = 0, \quad (2.35)$$

i.e.

$$(\sigma_0 p^0 - \sigma \cdot \mathbf{p})P\psi(x) = 0. \quad (2.36)$$

Going over to Fourier transform we have

$$(\sigma_0 p^0 + \sigma \cdot \mathbf{p})\psi_f(p^0, \mathbf{p}) = 0 \quad \text{and} \quad (\sigma_0 p^0 - \sigma \cdot \mathbf{p})P\psi_f(p^0, \mathbf{p}) = 0, \quad (2.37)$$

i.e. we must have

$$-\sigma \cdot \mathbf{p}\psi_f(p^0, \mathbf{p}) = P^{-1}\sigma \cdot \mathbf{p}P\psi_f(p^0, \mathbf{p}). \quad (2.38)$$

Taking  $\mathbf{p} \rightarrow -\mathbf{p}$  we have also

$$-\sigma \cdot \mathbf{p} \psi_f(p^0, -\mathbf{p}) = P^{-1} \sigma \cdot \mathbf{p} P \psi_f(p^0, -\mathbf{p}) . \quad (2.39)$$

$\psi_f(p^0, \mathbf{p})$  and  $\psi_f(p^0, -\mathbf{p})$  span the 2D space otherwise from  $\psi_f(p^0, -\mathbf{p}) = \alpha \psi_f(p^0, \mathbf{p})$  it follows

$$\sigma_0 p^0 \psi_f(p^0, \mathbf{p}) = 0 . \quad (2.40)$$

Thus

$$-\sigma \cdot \mathbf{p} = P^{-1} \sigma \cdot \mathbf{p} P, \quad \forall \mathbf{p} . \quad (2.41)$$

But such a matrix  $P$  does not exist as seen from

$$\{\sigma \cdot \mathbf{p}, a\sigma_0 + \sigma \cdot \mathbf{b}\} = 2a \sigma \cdot \mathbf{p} + 2\mathbf{b} \cdot \mathbf{p} . \quad (2.42)$$

Similar results hold for a  $\psi(x)$  which transforms as  $\psi'(x') = A\psi(x)$  and obeys  $\tilde{\sigma}_\mu p^\mu \psi(x) = 0$ . Weyl equation can be derived from the invariant action

$$S = \int d^4x \psi^\dagger(x) \sigma_\mu (-i\partial^\mu) \psi(x) . \quad (2.43)$$

Such an action is hermitean as seen by taking the hermitean conjugate and integrating by parts.

### 2.4.1 Majorana Mass

It is possible to give a mass to a two component field transforming according to the  $(1/2, 0)$  representation of the restricted Lorentz group as follows (similar considerations can be done for the  $(0, 1/2)$  representation). Let us consider the equation

$$\tilde{\sigma}_\mu p^\mu \phi(x) = -imc \sigma_2 \phi^*(x) . \quad (2.44)$$

We saw already that if  $\phi'(x') = A\phi(x)$  then  $\tilde{\sigma}_\mu p'^\mu \phi'(x') = \tilde{A} \tilde{\sigma}_\mu p^\mu \phi(x) \in (0, 1/2)$ . Thus we have simply to prove that  $\sigma_2 \phi^*(x) \in (0, 1/2)$ . In fact

$$\sigma_2 (A\phi(x))^* = \sigma_2 A^* \sigma_2 \phi^*(x) = \tilde{A} \sigma_2 \phi^*(x) . \quad (2.45)$$

Multiplying (2.44) on the left by  $\sigma_\mu p^\mu$  and using again (2.44) we obtain, taking into account that  $(p^\mu)^* = -p^\mu$ ,

$$-p^2 \phi(x) = m^2 c^2 \phi(x) \quad (2.46)$$

proving that both component of the spinor  $\phi(x)$  satisfy the massive Klein-Gordon equation.

Equation (2.44) can be derived by varying the action

$$S = \int d^4x \left[ \phi^+(x) i \tilde{\sigma}_\mu \partial^\mu \phi(x) + \frac{imc}{2} (\phi^T \sigma_2 \phi(x) - \phi^+ \sigma_2 \phi^*) \right]. \quad (2.47)$$

Again integrating by parts it is proved that the action  $S$  is hermitean.

The action (2.47) is invariant under the parity transformation

$$\phi'(x') = P\phi(x) = \sigma_2 \phi^*(x), \quad \text{with } x' = (x^0, -\mathbf{x}) \quad (2.48)$$

and thus also the equations of motion are invariant under such a transformation.

## 2.5 Dirac Equation

Consider again a  $\psi(x)$  which transforms according to

$$\psi'(x') = \tilde{A}\psi(x). \quad (2.49)$$

Posing

$$\sigma_\mu p^\mu \psi(x) = \phi(x) \quad (2.50)$$

we have that  $\phi(x)$  transforms like

$$\phi'(x') = A\phi(x). \quad (2.51)$$

In fact

$$A\sigma_\mu p^\mu A^+ \tilde{A}\psi(x) = A\phi(x), \quad (2.52)$$

i.e.

$$\sigma_\mu p'^\mu \psi'(x') = \phi'(x'). \quad (2.53)$$

We have now four components and we must supply other two equations, invariant under Lorentz transformations. These are given by

$$\tilde{\sigma}_\mu p^\mu \phi(x) = \kappa \psi(x), \quad (2.54)$$

which is invariant as

$$\tilde{A}\tilde{\sigma}_\mu p^\mu \tilde{A}^+ A\phi(x) = \kappa \tilde{A}\psi(x), \quad (2.55)$$



i.e.

$$\sigma_\mu p' \phi'(x') = \kappa \psi'(x') . \quad (2.56)$$

Multiplying now (2.50) on the left by  $\tilde{\sigma}_\mu p^\mu$  we have

$$- p_\mu p^\mu \psi(x) = \kappa \psi(x) , \quad (2.57)$$

i.e. the two equations,

$$\begin{aligned} \sigma_\mu p^\mu \psi(x) &= \phi(x) , \\ \tilde{\sigma}_\mu p^\mu \phi(x) &= \kappa \psi(x) , \end{aligned} \quad (2.58)$$

describe particles of mass  $m$  such that  $m^2 = \kappa$ . It is better to distribute in more symmetrical way the  $\kappa$

$$\begin{aligned} \sigma_\mu p^\mu \psi(x) &= m \phi(x) , \\ \tilde{\sigma}_\mu p^\mu \phi(x) &= m \psi(x) . \end{aligned} \quad (2.59)$$

Thus the

$$\Psi(x) = \begin{pmatrix} \psi(x) \\ \phi(x) \end{pmatrix} \quad (2.60)$$

transforms according to the  $(1/2, 0) \oplus (0, 1/2)$  representation. From  $\tilde{\sigma} = -\sigma$  it is immediately seen that the parity transformation is

$$\Psi'(x') = P\Psi(x), \quad \text{with } x' = (x^0, -\mathbf{x}) \quad (2.61)$$

and

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (2.62)$$

We can also write

$$\begin{aligned} \sigma^\mu p_\mu \psi(x) &= m \phi(x) , \\ \tilde{\sigma}^\mu p_\mu \phi(x) &= m \psi(x) , \end{aligned} \quad (2.63)$$

with  $\sigma^\mu = \eta^{\mu\nu} \sigma_\nu$  or

$$\left[ - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} p_0 + \begin{pmatrix} \sigma^k & 0 \\ 0 & -\sigma^k \end{pmatrix} p_k \right] \Psi(x) = m \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Psi(x) , \quad (2.64)$$

which can also be written as

$$(\gamma^\mu \partial_\mu + m) \Psi(x) = 0 , \quad (2.65)$$

with

$$\gamma^0 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} \quad \text{antihermitean} , \quad (2.66)$$

$$\gamma^j = \begin{pmatrix} 0 & -i\sigma^j \\ i\sigma^j & 0 \end{pmatrix} \quad \text{hermitean} . \quad (2.67)$$

Notice that defined

$$\Sigma_\mu = \begin{pmatrix} \tilde{\sigma}_\mu & 0 \\ 0 & \sigma_\mu \end{pmatrix} \quad (2.68)$$

and

$$\mathcal{A} = \begin{pmatrix} \tilde{A} & 0 \\ 0 & A \end{pmatrix} \quad (2.69)$$

we have obviously

$$\mathcal{A}\Sigma_\mu\mathcal{A}^+ = \Sigma_\nu\Lambda^\nu{}_\mu . \quad (2.70)$$

Taking into account that  $\gamma_\mu = i\Sigma_\mu P$  we have

$$\mathcal{A}\Sigma_\mu(iP)(-iP)\mathcal{A}^+(iP) = \Sigma_\nu(iP)\Lambda^\nu{}_\mu , \quad (2.71)$$

i.e.

$$\mathcal{A}\gamma_\mu\mathcal{A}^{-1} = \gamma_\nu\Lambda^\nu{}_\mu . \quad (2.72)$$

In addition it will be useful to define  $\gamma_5 = -i\gamma^0\gamma^1\gamma^2\gamma^3$  which in the same representation assumes the form

$$\gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{hermitean} . \quad (2.73)$$

## 2.6 Rarita-Schwinger Equation

In the theory of supergravity one needs the field theoretical description of spin 3/2 particles. These are described by a vector-spinor  $\psi^\mu$ . We recall that a 4-vector  $V^\mu$  has spin content  $0 \oplus 1$  because  $\partial_\mu V^\mu$  is a scalar. On the other hand a spinor describes a particle of spin 1/2. As a consequence the spin content of  $\psi^\mu$  is  $(0 \oplus 1) \otimes 1/2 = 1/2 \oplus 1/2 \oplus 3/2$  and thus we have two unwanted spin 1/2 particles. These can be eliminated by imposing the following supplementary conditions  $\gamma^\lambda\psi_\lambda = 0$  and  $\partial_\lambda\psi^\lambda = 0$ , which set to zero two spinors. For massive particles both conditions are contained in

the Rarita-Schwinger equation, which in addition contains the dynamical equations for the  $\psi_\lambda$ . Such equation is

$$\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\left(\partial_\rho + \frac{m}{2}\gamma_\rho\right)\psi_\sigma = 0, \quad (2.74)$$

where  $\varepsilon_{0123} = 1$  and  $\varepsilon^{0123} = -1$ . The only properties of the  $\gamma^\mu$  we shall use are the Clifford algebra

$$\{\gamma^\mu, \gamma^\lambda\} = 2\eta^{\mu\lambda} \quad (2.75)$$

and the ensuing relation  $\{\gamma^\lambda, \gamma_5\} = 0$ .

By taking the divergence of (2.74) we have, for  $m \neq 0$

$$\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\gamma_\rho\partial_\mu\psi_\sigma = 0. \quad (2.76)$$

But

$$\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\gamma_\rho = -i(\gamma^\mu\gamma^\sigma - \gamma^\sigma\gamma^\mu) \quad (2.77)$$

and thus

$$(\gamma^\mu\gamma^\sigma - \gamma^\sigma\gamma^\mu)\partial_\mu\psi_\sigma = 0. \quad (2.78)$$

Rewriting the above as

$$(2\gamma^\mu\gamma^\sigma - 2\eta^{\mu\sigma})\partial_\mu\psi_\sigma = 0 \quad (2.79)$$

we obtain

$$\gamma^\mu\partial_\mu(\gamma^\sigma\psi_\sigma) - \partial_\mu\psi^\mu = 0. \quad (2.80)$$

We contract now (2.74) with  $\gamma_\mu$  to obtain

$$\varepsilon^{\mu\nu\rho\sigma}\gamma_\mu\gamma_\nu\gamma_5\partial_\rho\psi_\sigma + \frac{m}{2}\varepsilon^{\mu\nu\rho\sigma}\gamma_\mu\gamma_\nu\gamma_5\gamma_\rho\psi_\sigma = 0. \quad (2.81)$$

The first term vanishes due to (2.76) and thus we have

$$0 = (\gamma^\rho\gamma^\sigma - \gamma^\sigma\gamma^\rho)\gamma_\rho\psi_\sigma = -2(\gamma^\sigma\gamma^\rho - \eta^{\sigma\rho})\gamma_\rho\psi_\sigma = -6\gamma^\sigma\psi_\sigma = 0. \quad (2.82)$$

Due to (2.80) it implies also  $\partial_\mu\psi^\mu = 0$ . Now we prove that each component of the Rarita-Schwinger field satisfies the Dirac equation. The term of the Rarita-Schwinger equation proportional to  $m$  can be written as

$$i\frac{m}{2}(\gamma^\mu\gamma^\sigma - \gamma^\sigma\gamma^\mu)\psi_\sigma = im(\gamma^\mu\gamma^\sigma - \eta^{\mu\sigma})\psi_\sigma = -im\psi^\mu. \quad (2.83)$$

With regard to the derivative term, it can be rewritten as

$$\begin{aligned} \varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\partial_\rho\psi_\sigma &= \varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\delta_\sigma^\lambda\partial_\rho\psi_\lambda = \\ &= \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5(\gamma^\lambda\gamma_\sigma + \gamma_\sigma\gamma^\lambda)\partial_\rho\psi_\lambda = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\gamma^\lambda\gamma_\sigma\partial_\rho\psi_\lambda. \end{aligned} \quad (2.84)$$

But the following equality holds

$$\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\gamma^\lambda\gamma_\sigma = i(\eta^{\rho\lambda}\gamma^\mu - \eta^{\mu\lambda}\gamma^\rho) , \quad (2.85)$$

which substituted in the previous expression gives  $-i\gamma^\mu\partial_\mu\psi^\mu$ . Summing the two terms we have

$$\gamma^\mu\partial_\mu\psi^\mu + m\psi^\mu = 0 . \quad (2.86)$$

We come now to the massless Rarita-Schwinger equation

$$\varepsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5\partial_\rho\psi_\sigma = 0 . \quad (2.87)$$

Such equation is invariant under the gauge transformation

$$\psi_\sigma \rightarrow \psi_\sigma + \partial_\sigma\eta \quad (2.88)$$

with  $\eta$  an arbitrary spinor. Contracting (2.87) with  $\gamma_\mu$  gives

$$\begin{aligned} 0 &= (\gamma^\rho\gamma^\sigma - \gamma^\sigma\gamma^\rho)\partial_\rho\psi_\sigma = 2(\gamma^\rho\gamma^\sigma - \eta^{\sigma\rho})\partial_\rho\psi_\sigma = \\ &= 2\gamma^\mu\partial_\mu(\gamma^\sigma\psi_\sigma) - 2\partial_\sigma\psi^\sigma . \end{aligned} \quad (2.89)$$

A useful gauge choice is  $\gamma^\sigma\psi_\sigma = 0$ , which induces  $\partial_\sigma\psi^\sigma = 0$ .

**Exercise 2.1:** If a charged particle is undeflected in passing through uniform crossed electric and magnetic fields  $E$  and  $B$  (mutually perpendicular, and both perpendicular to the direction of motion), what is its velocity? If we now turn off the electric field, and the particle moves in an arc of radius  $R$ , what is its charge-to-mass ratio?

**Exercise 2.2:** Find expression of current for the complex scalar field.

**Exercise 2.3:** Construct the normalized spinors representing an electron of momentum  $p$  with helicity  $\pm 1$ .

**Exercise 2.4:** Construct the Hamiltonian for the Dirac equation.

**Exercise 2.5:** Find the commutator of the Dirac Hamiltonian with the orbital angular momentum.

**Exercise 2.6:** Find the commutator of the Dirac Hamiltonian with the spin angular momentum.

**Exercise 2.7:** Show that the spinor representing an electron at rest is an eigenstate of the parity operator.

## Chapter 3

# Forces and Particles

### 3.1 The Fundamental Forces

The two forces most familiar to people, *Gravity* and *Electromagnetism*, are only half of the four fundamental forces in our universe (that we know of).

We can think about the third force by considering a compact nucleus which we know to be made of protons and neutrons. We know that the protons should repel each other because of their like charge. But the nuclei of atoms somehow hold together, which is evidence for some stronger force that causes these particles to attract. This force, which allows atomic nuclei to remain stable, is called the *Strong* force (The  $SU(3)$  color force). Just as electrically-charged particles are subject to the electromagnetic force, some particles have a property similar to charge, called *Color*, and are subject to the strong force. The field theory that describes this is called *Quantum Chromodynamics* (QCD) and was first proposed in 1965 by Han, Nambu, and Greenberg. This theory predicts the existence of the gluon, which is the mediator of the strong force between two matter particles.

The fourth force is the one we have the least familiarity with. It is responsible for certain types of radioactive decays; for example, permitting a proton to turn into a neutron and vice versa. It is called the *Weak* force (The  $SU(2)$  part that is left over when  $SU(2) \otimes U(1)$  is broken).

In the 1960's, Sheldon Glashow, Abdus Salam, and Steven Weinberg in-

independently developed a theory that unified the electromagnetic and weak force. At sufficiently high energies it is observed that the difference between these two separate forces is negligible and that they act together as the *Electro-Weak* (EW) force (The unbroken  $SU(2) \otimes U(1)$  force). For processes at lower energy scales, the symmetry between the electromagnetic and the weak force is broken and we observe two different forces with different properties. EW theory predicts four force-carrier particles, corresponding to the 3 generators of  $SU(2)$  and the 1 in  $U(1)$ , that mediate the force between matter particles. Two of these particles are uncharged, while two carry charge. The mediating particle for electromagnetism is the photon, and those for the weak force are  $W^\pm$  (with  $\pm 1$  electron charges) and  $Z^0$  (neutral) bosons.

The electromagnetic, weak, and strong forces form what is called the *Standard Model* (SM) of Particle Physics. The SM is an incomplete theory in the sense that it fails to describe gravitation, the force that acts on matter.

## 3.2 Fundamental Particles

In the V century B.C., a Greek named Empedocles took the ideas of several others before him and combined them to say that matter is made up of earth, wind, fire, and water, and that there are two forces, Love and Strife, that govern the way they grow and act. More scientifically, he was saying that matter is made of smaller substances that interact with each other through repulsion and attraction.

Democritus, a contemporary of Empedocles, went a step further to say that all matter is made of indestructible fundamental particles. He called these particles atoms, meaning 'indivisible'. Of course, our modern use of the word is different. At their discovery, it was thought that different elements were the indivisible particles sought for, so the name atom seemed appropriate.

In the last century, experimenters were surprised as they discovered new particle after new particle. It seemed disorganized and overwhelming that there could be so many elementary objects. Eventually, however, the properties of these particles became better understood and it was found that there really is just a small, finite set of fundamental particles, some of which can be grouped together to make up larger objects.

One property of the "zoo" of discovered particles that helps in our organizing

them is their intrinsic spin. Any particle, elementary or composite, that is of half-integer spin is a *Fermion*. Those with integer spin are *Bosons*. The spins govern the statistics of a set of such particles, so fermions and bosons may also be defined according to the statistics they obey.

Namely, fermions obey *Fermi-Dirac* statistics and therefore also obey the *Pauli Exclusion Principle*. This means that no two identical fermions can be found in the same quantum state at the same time. Furthermore, to accurately display this behavior it is found that the wave function of a system with fermions must be antisymmetric; swapping any two like fermions causes a change in sign of the overall wave function.

Bosons on the other hand obey *Bose-Einstein* statistics; any number of the same type of particle can be in the same state at the same time. In contrast to fermions, the wavefunction of a system of bosons is symmetric.

The elementary particles are those that are considered fundamental, or in other words, are not composed of smaller particles. They can be divided into two groups: matter particles and non-matter particles. The elementary matter particles all have half-integer spin (so are fermions) and the elementary non-matter particles all have integer spin (so are bosons).

### 3.2.1 Elementary Fermions

The elementary fermions are the building blocks of all other matter. For example, the proton and neutron are made up of different combinations of three elementary quarks. Electrons, which are also elementary, cloud around the protons and neutrons, and when all three group together in a particular way, an atom is formed. Less familiar examples include those that are unstable, such as the muon, which decay fairly quickly.

For every elementary particle, there is also a corresponding particle with the same mass but of different charge and magnetic moment. Generally the name of such a particle is the same as the corresponding 'normal' matter particle, but with the prefix 'anti' in front of it (e.g. antiquark, antilepton, etc.). Whenever we discuss matter and its properties, it is implied that the antimatter counterparts have similar properties.

Now we further divide the elementary fermions into two groups, quarks and leptons. A convenient way to distinguish these two sets is by whether or not

they interact via the strong force: quarks may interact via the strong force, while leptons do not.

**Quarks:** Experiments involving high energy collisions of electrons and protons led Murray Gell-Mann to suggest in 1964 that protons and neutrons are actually composite particles, made of three point-like, spin-1/2 particles whose charges are either  $-1/3$  or  $+2/3$  units of electron charge. He called these particles *Quarks*. Through further experiments it has been found that there are six flavors of quarks total, grouped into three generations with the first generation containing the up and down quarks, the second generation containing the more massive charm and strange quarks, and the third generation containing the even more massive top and bottom quarks.

As electrically charged particles are subject to the electromagnetic force, quarks have a property similar to charge, called color, and any colored particle is subject to the strong force. It is found that there are three different types of colors: (defined as) red, green, and blue (plus three more for anti-quarks: antired, antigreen, and antiblue). Quarks are grouped together to make composite particles that are colorless (the color charges cancel out), which is why the concept of color was only discovered after quarks themselves were found. The addition of color to the quark model also ensures that any quarks contained in a composite particle will not violate the exclusion principle since each has a different color. Again, QCD is the field theory that describes these properties.

Another interesting feature of quarks is that they are never found alone, but rather always inside of a composite particle. This phenomenon is called *Confinement*. It is more a property of the strong force, which increases in strength as two colored particles are pulled away from each other, just as would happen when the ends of a piece of elastic are pulled apart. We can consider reaching a distance between the two quarks where there is sufficient potential energy built up that it can be converted to matter, creating a quark-antiquark pair. The pair will separate and the resulting particles will recombine with the original quarks. As this process repeats, and more quark-antiquark pairs are created, the end result in the whole process is a multiplication of the number of quarks and of the number of composite particles. In the opposite extreme, as two quarks get closer together, the strong force between them becomes weaker until the quarks move around freely and more independently. This is called *Asymptotic Freedom*.

Quarks also interact with other particles via the weak force, which is the



only force that can cause a change of flavor (changing an up into a down, for example). When this happens, a quark either turns into a heavier quark by absorbing a  $W$ -boson, or it emits a  $W$ -boson and then decays to a lighter quark. Beta decay, a common radioactive process, is caused by this mechanism. Instead of just thinking of beta decay as a neutron in the nucleus of an atom decaying, or splitting, into a proton, electron, and antineutrino, we can go a step further with our understanding of quarks subject to the weak force. We add that, really, it is one of the down quarks in the neutron that emits a  $W^-$  boson and then decays to the lighter up quark, keeping charge conserved in the process. Other conserved quantities are momentum, energy, quark number, lepton number, and (approximately) lepton generation number. The neutron, which used to have one up and two down quarks, now has one down and two up quarks, which is the composition of a proton. The electron and antineutrino are created from the decay of the  $W^-$  boson.

**Leptons:** *Leptons* interact with other matter via the electromagnetic, the weak, and gravitational forces, but not through the strong force. This means that leptons carry  $SU(2) \otimes U(1)$  terms in their covariant derivatives, but not  $SU(3)$  terms. There are three charged leptons, grouped, like the quarks, into three different generations based on their masses. This is equivalent to the statement that there are three copies of the SM Gauge Group. The electron is the lightest of the charged leptons, then the muon, and the tau. There are also three neutral leptons, called neutrinos, one type for each of the charged leptons: the electron neutrino, the muon neutrino, and the tau neutrino.

Some quantities in lepton events are found to be conserved. These conservation laws can all be derived from some rules, though they are typically treated separately because they are extremely useful when talking about specific interactions. If we define lepton number as the number of leptons minus the number of antileptons, then lepton number is constant in all interactions. Additionally, the lepton number within each generation is also approximately conserved. For example, the number of electrons and electron neutrinos minus the number of antielectrons and electron antineutrinos is found to be constant in most particle reactions.

An interesting exception is in neutrino oscillations, where a neutrino changes lepton flavors as it travels. For example, we can take a measurement and observe an electron neutrino, even though it was known to have been created as a muon neutrino. These oscillations of flavor only occur if neutrinos have mass (even just very small mass), so the fact that the SM currently predicts

them to be massless demonstrates that there are some parameters in the theory that need to be adjusted.

### 3.2.2 Elementary Bosons

Throughout the development of the SM it was found that some particles play a different role than the 'matter' particles that make up the stuff of the universe. Both the gauge bosons and the Higgs boson fall into this group.

**Gauge Bosons:** In QFT, the Lagrangian can be made invariant under a local gauge transformation by the addition of a vector field called a gauge field. The quanta of this field is called a *Gauge Boson*. There are three types of gauge bosons described by the SM (i.e. there are three gauge groups, each with their own set of generators). They are the photon, which carries the electromagnetic force, the  $W^\pm$  and  $Z$  bosons, which carry the weak force, and the gluons which carry the strong force. Each of these bosons have been experimentally detected.

Evidence for the photon first came in 1905 when Einstein proposed an explanation of the photo-electric effect, that light was quantized into energy packets. Confirmation of the  $W^\pm$  and  $Z^0$  bosons came in 1983 through proton-proton collisions at the CERN.

The gluons were first experimentally observed in 1979 in the electron-positron collider at the German Electron Synchrotron (DESY) in Hamburg. Further experiments have demonstrated that the gluons have eight different color states and that, because they interact via the strong force, they have properties similar to quarks, such as confinement.

Taking into account their possible charge or color, we find that there are 12 gauge bosons in all, one for the electromagnetic force, three for the weak force, and eight for the strong force.

**Higgs Boson:** The Higgs boson was discovered in 2012 at CERN. It is the only elementary boson that is not a gauge boson. Rather, it is the carrier of the scalar field from which other particles acquire mass. The existence of the Higgs explains why some particles have mass and others do not.

Let us provide the Table 3.2.2 which should help you see this all more clearly, the relative strengths have been normalized to unity for the strong force.

Interaction	Gauge Boson	Acts On	Strength	Range
Strong	Gluon	Hadrons	1	$10^{-15}$ m
Electro-Magnet.	Photon	Electric Charges	$10^{-2}$	$\infty (1/r^2)$
Weak	$W^\pm, Z^0$	Lept. and Hadr.	$10^{-5}$	$10^{-18}$ m
Gravity	Graviton	Mass	$10^{-39}$	$\infty (1/r^2)$

Table 3.1: Fundamental forces of nature.

### 3.3 Periodic Table of Particle Physics

As things stand today, the periodic table of the SM is complete. One part of this periodic table are the spin-1/2 matter particles: the quarks and the leptons and their anti-particles. The Table 3.2 summarises the details of the currently available information on all the matter fermions.

Quarks	Leptons
$2/3 \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$ $-1/3$	$0 \begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$ $-1$
$2/3 \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$ $-1/3$	
$2/3 \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$ $-1/3$	
$M_u = 2 \text{ MeV}$ $M_d = 5 \text{ MeV}$ $M_c = 1,300 \text{ MeV}$ $M_s = 100 \text{ MeV}$ $M_t = 173.000 \text{ MeV}$ $M_b = 4.200 \text{ MeV}$	$M_{\nu_1} = 0 - 0.13 \times 10^{-6} \text{ MeV}$ $M_e = 0.511 \text{ MeV}$ $M_{\nu_2} = 0.009 - 0.13 \times 10^{-6} \text{ MeV}$ $M_\mu = 106 \text{ MeV}$ $M_{\nu_3} = 0.04 - 0.14 \times 10^{-6} \text{ MeV}$ $M_\tau = 1.777 \text{ MeV}$

Table 3.2: Elementary fermions of the SM, all of spin 1/2. The three quark colours are indicated explicitly, while leptons are colourless. Electric charges in units of the positron charge, are displayed on the left side. The anti-particles form a similar table with opposite charges.

Of course, a gauge field theoretic description of the interactions among these elementary particles needs in the SM particle spectrum, also the gauge bosons which would be the carrier of the various interactions. This leads

to the second set of members of the 'periodic table' of particle physics, viz. the spin-1 gauge bosons: the photon,  $W$  and  $Z$  bosons and gluons.

As we will discuss in detail later, gauge invariance, which guarantees the renormalisability of this theory, would require that all of the gauge bosons should be massless. Not only that, the same invariance would require the matter fermions also to be massless. However, other than the gluon and the photon all the other members of this periodic table are patently massive. In fact, it is the mechanism of Spontaneous Symmetry Breaking (SSB), which allows these particles to have non-zero masses and helps keep the theory still consistent with gauge invariance. SSB of the EW gauge symmetry via the Higgs mechanism, is the key ingredient of renormalisable gauge theories of the EW interaction. This requires existence of yet another member of the periodic table, which is the Higgs boson. This too has been included in the list of the SM bosons in Table 3.3.

Electromagnetic and weak (Spin 1)	Strong (Spin 1)	Higgs (Spin 0)
$\gamma$ (photon) $W^\pm, Z$ (weak bosons)	$g$ (gluons)	$h$ (Higgs)
$M_\gamma = 0$ $Q_\gamma = 0$ $M_W = 80.404 \text{ GeV}$ $Q_W = \pm 1$ $M_Z = 90.1876 \text{ GeV}$ $Q_Z = 0$	$M_g = 0$ $Q_g = 0$	$M_h = 125.4 \pm \text{ GeV}$ $Q_h = 0$

Table 3.3: Elementary bosons of the SM. There are no separate anti-particles:  $W^-$  is the anti-particle of  $W^+$  and the rest are neutral.  $Q$  indicates the electromagnetic charge of the boson in units of positron charge.

### 3.4 Composite Particles

Examples of composite particles are hadrons, nuclei, atoms and molecules. The latter three are well known and will not be described here.

Hadrons are made up of bound quarks and interact via the strong force. They can be either fermions or bosons, depending on the number of quarks that

make them up. An odd number of bound quarks create a spin-1/2 or spin-3/2 hadron, which is called a baryon, and an even number of quarks create spin-0 or spin-1 hadrons, called mesons. Experimentally, only combinations of three quarks or two quarks have been found, so the terms baryon and meson often just refer to three or two bound quarks, respectively.

You can understand why mesons and baryons have the spin that they do by considering how many spin-1/2 quarks compose them. A meson has two quarks, and therefore the total spin of a meson is the sum of an even number of half-integer spin particles, which will be integer spin. And because there are only two of them, it is either spin 0 or 1. Baryons, on the other hand, will have a linear combination of three particles with half-integer spin, which will of course be half-integer: 1/2 or 3/2.

The most well-known examples of baryons are protons and neutrons. Protons are made of two up quarks and one down quark, or  $|uud\rangle$ , and neutrons are made of two down and one up, or  $|udd\rangle$ . The baryons are made of 'normal' quarks only and their antimatter counterparts are made of the corresponding antiquarks.

The mesons are made of a quark and an antiquark pair, though not necessarily of the same generation. Examples are  $\pi^+$ ,  $|u\bar{d}\rangle$  and  $K^+$ ,  $|u\bar{s}\rangle$ .

One of the reasons for the "zoo" of particles discovered in the past century is because of the numerous possible combinations of six quarks put into a three-quark or two-quark hadron. Additionally, each of these combinations can be in different quantum mechanical states, thereby displaying different properties. For example, a rho meson  $\rho$  has the same combination of quarks as a pion  $\pi$ , but the  $\rho$  is spin-1 whereas the pion is spin-0.

**Exercise 3.1:** Calculate the ratio of the gravitational attraction to the electrical repulsion between two stationary electrons (Do you need to know how far apart they are?).

**Exercise 3.2:** How many different meson combinations can you make with 1, 2, 3, 4, 5, or 6 different quark flavors? What's the general formula for  $n$  flavors?

**Exercise 3.3:** How many different baryon combinations can you make with 1, 2, 3, 4, 5, or 6 different quark flavors? What's the general formula for  $n$  flavors?

**Exercise 3.4:** Using four quarks ( $u$ ,  $d$ ,  $s$ , and  $c$ ), construct a table of all the possible baryon species. How many combinations carry a charm of +1? How many carry charm +2, and +3?

**Exercise 3.5:** Using four quarks ( $u$ ,  $d$ ,  $s$ , and  $c$ ), construct a table of all the possible meson species. How many combinations carry a charms of +1, +2 and +3?

**Exercise 3.6:** Suppose that all quarks and leptons are composed of two more elementary constituents - Preons:  $c$  (with charge  $-1/3$ ) and  $n$  (with charge zero) - and their respective antiparticles. You're allowed to combine them in groups of three preons or three anti-preons ( $ccn$ , for example, or  $nnn$ ). Construct all of the eight quarks and leptons in the first generation in this manner. (The other generations are supposed to be excited states.) Notice that each of the quark states admits three possible permutations ( $ccn$ ,  $cnc$ ,  $ncc$ , for example) - these correspond to the three colors. Mediators can be constructed from three preons plus three anti-preons.  $W^\pm$ ,  $Z^0$ , and  $\gamma$  involve three preons and three anti-preons ( $W^- = ccn\bar{n}\bar{n}$ , for instance). Construct  $W^+$ ,  $Z^0$ , and  $\gamma$  in this way. Gluons involve mixed combinations ( $ccn\bar{c}\bar{n}$ , for instance). How many possibilities are there in all? Can you think of a way to reduce this down to eight?

**Exercise 3.7:** A quark and an antiquark are bound together, in a state of zero orbital angular momentum, to form a meson. What are the possible values of the meson's spin?

**Exercise 3.8:** Suppose you combine three quarks in a state of zero orbital angular momentum. What are the possible spins of the resulting baryon?

## Part II

### Lecture – Particle Reactions





## Chapter 4

# Elements of the Scattering Theory

### 4.1 Kinematics of the Minkowski Spacetime

In particle physics, the particles are treated relativistically, meaning  $E \approx pc \gg mc^2$  and thus special theory of relativity becomes an mathematical tool in describing the particle kinematics.

Introducing the position 4-vector,  $x^\mu$ , one can show that the 4-dimensional length element,

$$s \equiv (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2, \quad (4.1)$$

is Lorentz Invariant. This is like  $r^2 = x^2 + y^2 + z^2$  being invariant under spatial rotation. The quantity  $s$  could be written in the form of a sum:  $s = g_{\mu\nu}x^\mu x^\nu$ , where the metric  $g_{\mu\nu}$  is defined as

$$g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \quad (4.2)$$

Let us define covariant 4-vector  $x_\mu$  (index down),  $x_\mu = g_{\mu\nu}x^\nu$ , and contravariant 4-vector  $x^\mu$  (index up). Then

$$s = x_\mu x^\mu = x^\mu x_\mu. \quad (4.3)$$

To each contravariant 4-vector  $a^\mu$ , a covariant 4-vector could be assigned and vice-versa.

$$a^\mu = g^{\mu\nu} a_\nu, \quad a_\mu = g_{\mu\nu} a^\nu, \quad (4.4)$$

where  $g^{\mu\nu}$  are the elements in  $g^{-1}$ . Since  $g^{-1} = g$ , we have

$$g^{\mu\nu} = g_{\mu\nu}. \quad (4.5)$$

Given any two 4-vectors,  $a^\mu$  and  $b^\mu$ ,

$$a^\mu b_\mu = a_\mu b^\mu = a^0 b^0 - a^1 b^1 - a^2 b^2 - a^3 b^3 \quad (4.6)$$

is Lorentz invariant. The vector  $a^\mu$  is called *time-like*, if  $a^2 > 0$ , *space-like* if  $a^2 < 0$  and *light-like* if  $a^2 = 0$ .

The 3-velocity of a particle is given by  $\vec{v} = d\vec{x}/dt$ , where  $d\vec{x}$  is the distance travelled in the laboratory frame and  $dt$  is the time measured in the same frame. Proper velocity of the particle is given by  $\vec{\eta} = d\vec{x}/d\tau$ , where  $d\vec{x}$  is the distance travelled in the laboratory frame and  $d\tau \equiv dt\sqrt{1-\beta^2}$  ( $\beta \equiv |v/c|$ ) is the proper time. Now

$$\vec{\eta} = \frac{d\vec{x}}{d\tau} = \frac{d\vec{x}}{dt} \frac{dt}{d\tau} = \vec{v}\gamma \quad \Rightarrow \quad \vec{\eta} = \gamma\vec{v}. \quad (4.7)$$

It is easy to work with the proper velocity,  $\vec{\eta}$ , as only  $d\vec{x}$  transforms under Lorentz transformation. Furthermore,

$$\eta^0 = \frac{dx^0}{d\tau} = \frac{d(ct)}{dt/\gamma} = \gamma c, \quad (4.8)$$

hence

$$\eta^\mu = \gamma(c, v_x, v_y, v_z). \quad (4.9)$$

This is called the *proper velocity 4-vector*. Remember that the spatial component brings up the negative sign for covariant tensor. Now

$$\eta^\mu \eta_\mu = \gamma^2 (c^2 - v_x^2 - v_y^2 - v_z^2) = \gamma^2 c^2 \left(1 - \frac{v^2}{c^2}\right) = c^2, \quad (4.10)$$

which is Lorentz Invariant. This also proves that 4-vector scalar product is Lorentz invariant.

Successive Lorentz boost in the same direction is represented by a single boost, where the transformation velocity is given by

$$\beta'' = \frac{\beta + \beta'}{1 + \beta\beta'}. \quad (4.11)$$

The velocity,  $\beta$ , is not an additive quantity, i.e. non-linear in successive transformation. Here comes the need of "*Rapidity*",  $y$ , to circumvent this drawback, by defining

$$\beta = \tanh(y) , \quad y = \frac{1}{2} \ln \left( \frac{1 + \beta}{1 - \beta} \right) . \quad (4.12)$$

One can show that rapidity is an additive quantity, i.e.

$$y'' = y + y' . \quad (4.13)$$

Using the rapidity, a Lorentz transformation with finite  $\eta$ , can be decomposed into  $N$  successive transformations with rapidity  $\Delta y = \eta/N$ . Solving  $\beta$ ,  $\gamma$  in terms of  $y$ , we have

$$\beta = \tanh(y) , \quad \gamma = \cosh(y) , \quad \beta\gamma = \sinh(y) . \quad (4.14)$$

Using  $y$  Lorentz boost can be written as:

$$\begin{aligned} x^{0'} &= \cosh(y) x^0 + \sinh(y) x^1 , \\ x^{1'} &= \sinh(y) x^0 + \cosh(y) x^1 . \end{aligned} \quad (4.15)$$

In high-energy collider experiments, the secondary particles which are produced from the interaction, are boosted in the  $z$ -direction (along the beam axis). The boosted angular distribution is better expressed as rapidity distribution. At high-energies, each particle has  $E \sim pc$ , and the parallel momentum  $p_{\parallel} = p \cos(\theta)$ , and its rapidity is approximated by so-called pseudo-rapidity:

$$\eta' = \frac{1}{2} \ln \left( \frac{1 + \beta_{\parallel}}{1 - \beta_{\parallel}} \right) = \frac{1}{2} \ln \left( \frac{E + p_{\parallel}c}{E - p_{\parallel}c} \right) \sim -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right] . \quad (4.16)$$

This fact is taken into account in designing detectors, which are divided into modules that span the same solid angle in the  $\eta - \phi$  (azimuthal angle) plane.

We know *momentum* = *mass*  $\times$  *velocity*. And velocity can be "ordinary velocity" or "proper velocity". Classically, both are equal (non-relativistic limit). If  $\vec{p} = m\vec{v}$ , the conservation of momentum is inconsistent with the principle of relativity. In relativity, momentum is the product of mass and proper velocity,

$$\vec{p} \equiv m\vec{\eta} = \gamma m\vec{v} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} . \quad (4.17)$$

In 4-dimensional formalism this is the spatial component of  $p^\mu$ ,

$$p^\mu = m\eta^\mu . \quad (4.18)$$

The temporal component,  $p^0 = \gamma mc$ , constitutes the relativistic energy:

$$E \equiv \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} . \quad (4.19)$$

Hence, the energy-momentum 4-vector:

$$p^\mu = \left( \frac{E}{c}, p_x, p_y, p_z \right) . \quad (4.20)$$

Now

$$p^\mu p_\mu = \frac{E^2}{c^2} - \vec{p}^2 = (m\eta^\mu)(m\eta_\mu) = m^2(\eta^\mu\eta_\mu) = m^2c^2 . \quad (4.21)$$

For ordinary massive particle  $p^2 = m^2 > 0$ , for massless particles like photons, gravitons etc.  $p^2 = m^2 = 0$ , for tachyons or virtual particles  $p^2 < 0$ , and  $p^\mu = 0$  corresponds to the vacuum state.

Remember that the relativistic equations  $\vec{p} = \gamma m\vec{v}$  and  $E = \gamma m$  do not hold good for massless particles and  $m = 0$  is allowed only if the particle travels with the speed of light. For massless particles,  $v = c$  and  $E = |\vec{p}|c$ .

## 4.2 Collider vs Fixed Target Experiment

### 4.2.1 Mandelstam Variables

Mandelstam variables  $s$ ,  $t$  and  $u$  are often used in scattering calculations. They are defined (for  $1 + 2 \rightarrow 3 + 4$  scattering) as

$$s = (p_1 + p_2)^2, \quad t = (p_1 - p_3)^2, \quad u = (p_1 - p_4)^2 . \quad (4.22)$$

This means that only two Mandelstam variables are independent. Their main advantage is that they are Lorentz invariant which renders them convenient for Feynman amplitude calculations. Only at the end we can exchange them for 'experimenter's' variables  $E$  and  $\theta$ .

### 4.2.2 Symmetric Collisions ( $A + A$ )

Consider the collision of two particles. The total 4-momentum of the system is a conserved quantity in the collision.

In *Laboratory System* (LS), the particle with momentum  $\mathbf{p}_1$ , energy  $E_1$  and mass  $m_1$  collides with a particle of mass  $m_2$  at rest. The 4-momenta of the particles are:

$$p_1 = (E_1, \mathbf{p}_1) , \quad p_2 = (m_2, \mathbf{0}) . \quad (4.23)$$

Then,

$$p_1 p_2 = E_1 m_2 , \quad (4.24)$$

and

$$p_\mu p^\mu = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2E_1 m_2 , \quad (4.25)$$

hence

$$E_{cm} = \sqrt{s} = \sqrt{m_1^2 + m_2^2 + 2E_{proj} m_2} , \quad (4.26)$$

where  $E_1 = E_{proj}$ , the projectile energy in LS.

In *Center of Mass* (CM) system, the momenta of both the particles are equal and opposite, the 4-momenta are:

$$p_1^* = (E_1^*, \mathbf{p}_1^*) , \quad p_2^* = (E_2^*, -\mathbf{p}_1^*) . \quad (4.27)$$

Then,

$$p_\mu p^\mu = (p_1 + p_2)^2 = (E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2 = E_{cm}^2 \equiv s , \quad (4.28)$$

$\sqrt{s}$  is the total energy in the CM, which is the invariant mass of the CM. For the given system, the two particles  $p_1$  and  $p_2$  are equivalent to one single particle with 4-momentum  $p_1 + p_2$  and mass  $E_{cm}$ . Generalizing this, one can consider these individual particles to represent a system of particles.

Further, we know

$$\mathbf{p} = m\vec{v}\gamma , \quad \text{and} \quad E = m\gamma , \quad (4.29)$$

$\gamma = 1/\sqrt{1 - \beta^2}$ . So, one obtains:

$$\beta_{cm} = \frac{\mathbf{p}_1 + \mathbf{p}_2}{E_1 + E_2} , \quad (4.30)$$

which is the velocity of the CM seen from the LS. It is evident here that the CM frame with an invariant mass  $\sqrt{s}$  moves in the laboratory in the direction of  $\mathbf{p}_1$  with a velocity corresponding to the Lorentz factor,

$$\begin{aligned}\gamma_{cm} &= \frac{1}{\sqrt{1-\beta^2}} = \frac{E_1 + E_2}{\sqrt{(E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2}} = \\ &= \frac{E_1 + m_2}{\sqrt{s}}, \quad \Rightarrow \quad \sqrt{s} = \frac{E_{lab}}{\gamma_{cm}},\end{aligned}\quad (4.31)$$

this is because  $E = \gamma m$  and the rapidity

$$y_{cm} = \cosh^{-1} \gamma_{cm} . \quad (4.32)$$

The center of mass, or center of momentum frame, is at rest and the total momentum is zero. This makes it a suitable choice for solving kinematics problems.

We know that for a collider with head-on collision ( $\theta = 180^\circ$ )

$$E_{cm}^2 = m_1^2 + m_2^2 + 2(E_1 \cdot E_2 + |\mathbf{p}_1| |\mathbf{p}_2|) . \quad (4.33)$$

For relativistic collisions,  $m_1, m_2 \ll E_1, E_2$ ,

$$E_{cm}^2 \simeq 4E_1 E_2 . \quad (4.34)$$

For two beams crossing at an angle  $\theta$ ,

$$E_{cm}^2 = 2E_1 E_2 (1 + \cos \theta) . \quad (4.35)$$

The CM energy available in a collider with equal energies ( $E$ ) for new particle production rises linearly with  $E$ , i.e.

$$E_{cm} \simeq 2E . \quad (4.36)$$

For a fixed-target experiment the CM energy rises as the square root of the incident energy:

$$E_{cm} \simeq \sqrt{2m_2 E_1} . \quad (4.37)$$

Hence the highest energy available for new particle production is achieved at collider experiments.

Most of the times the energy of the collision is expressed in terms of nucleon-nucleon center of mass energy. In the nucleon-nucleon CM frame, two nuclei

approach each other with the same boost factor  $\gamma$ . The nucleon-nucleon CM is denoted by  $\sqrt{s_{NN}}$  and is related to the total CM energy by

$$E_{cm} = A\sqrt{s_{NN}} . \quad (4.38)$$

This is for a symmetric collision with number of nucleons in each nuclei as  $A$ . The colliding nucleons approach each other with energy  $\sqrt{s_{NN}}/2$  and with equal and opposite momenta. The rapidity of the nucleon-nucleon center of mass is  $y_{NN} = 0$  and taking  $m_1 = m_2 = m_p$ , the projectile and target nucleons are at equal and opposite rapidities,

$$y_{proj} = -y_{target} = \cosh^{-1} \frac{\sqrt{s_{NN}}}{2m_p} = y_{beam} . \quad (4.39)$$

For the Lorentz Factor we have:

$$\gamma = \frac{E}{M} = \frac{E_{cm}}{2A m_p} = \frac{A \sqrt{s_{NN}}}{2A m_p} = \frac{\sqrt{s_{NN}}}{2 m_p} = \frac{E_{beam}^{cm}}{m_p} ,$$

where  $E$  and  $M$  are Energy and Mass in CM respectively. Assuming mass of a proton,  $m_p \sim 1$  GeV, the Lorentz factor is of the order of beam energy in CM for a symmetric collision.

### 4.2.3 Asymmetric Collisions ( $A + B$ )

The CM energy of a collision of two different systems with charge  $Z_1$ ,  $Z_2$  and atomic numbers  $A_1$ ,  $A_2$  with  $Z = A = 1$ , for a proton is

$$\sqrt{s_{NN}} \simeq 2\sqrt{s_{pp}} + \sqrt{s_{pp}} \sqrt{\frac{Z_1 Z_2}{A_1 A_2}} , \quad (4.40)$$

where sub-index  $NN$  refers to the energy per nucleon inside the colliding nucleus and  $\sqrt{s_{pp}}$  is the corresponding energy in  $pp$  collisions. The rapidity shift in non-symmetric systems is given by

$$\Delta y \simeq \frac{1}{2} \ln \left[ \frac{Z_1 A_2}{Z_2 A_1} \right] . \quad (4.41)$$

This is due to the fact that the center-of-mass frame of the  $pA$  collision doesn't coincide with the laboratory center-of-mass frame. The rapidity shift in  $p + Pb$  collision is

$$\Delta y \simeq \frac{1}{2} \ln \left[ \frac{Z_1 A_2}{Z_2 A_1} \right] = \frac{1}{2} \ln \left[ \frac{82 \times 1}{1 \times 208} \right] = -0.465 . \quad (4.42)$$

This rapidity shift need to be taken into account for the comparison with  $Pb + Pb$  data.

At LHC, the maximum proton beam energy is 7 TeV, while the maximum  $Pb$  beam energy is 2.75 TeV,  $E_{cm} = 8.775$  TeV. The difference in available energy is due to the charge-to-mass ratio,  $Z/A$ . More is the number of neutrons in the nucleus, difficult it is to accelerate to higher energies. Because of different energies, the two beams will also not have the same rapidity. For the proton beam  $y_p = 9.61$  and for the  $Pb$  beam it is  $y_{Pb} = 8.67$ . Thus the center of the collision is shifted away from  $y_{cm} = 0$  by

$$\Delta y_{cm} = \frac{1}{2}(y_p - y_{Pb}) = 0.47 . \quad (4.43)$$

**Exercise 4.1:** Prove that  $s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2$ .

**Exercise 4.2:** A pion at rest decays into a muon plus a neutrino. What is the speed of the muon?

**Exercise 4.3:** Suppose two identical particles, with equal velocities, collide head-on. What is the kinetic energy of one in the rest system of the other?

**Exercise 4.4:** In a fixed-target  $pp$  experiment, what proton energy would be required to achieve the same CM energy as the LHC, which is operating at 14 TeV.

**Exercise 4.5:** Show that the process  $\gamma \rightarrow e^+e^-$  cannot occur in the vacuum.

**Exercise 4.6:** For the decay  $a \rightarrow 1 + 2$ , express the mass of the particle  $a$  by the velocities of the daughter particles and by the angle between them.

**Exercise 4.7:** In the LS,  $p$  with total energy  $E$  collides with  $p$  at rest. Find the minimum  $E$  such that process  $p + p \rightarrow 2p + 2\bar{p}$  is kinematically allowed.

**Exercise 4.8:** At the HERA collider, 27.5 GeV electrons were collided head-on with 820 GeV protons. Calculate the CM energy.

**Exercise 4.9:** Find the energy of the photon after the Compton scattering from an electron at rest.

**Exercise 4.10:** Suppose we interpret the electron as a classical solid sphere spinning with angular momentum  $\hbar/2$ . What is the speed of a point on its "equator"? It is known that electrons radius is less than  $10^{-16}$  cm.



## Chapter 5

# Golden Rules for Particle Reactions

### 5.1 Free Field Solutions

Processes in particle physics are mostly calculated in the framework of the quantum field theory. However, predictions of quantum field theory pertaining to the elementary particle interactions can often be calculated using a relatively simple "recipe" – *Feynman diagrams*.

Before we turn to describing the method of Feynman diagrams, let us just specify quantum fields that take part in the elementary particle physics interactions. All these are free fields, and interactions are treated as their perturbations. Each particle type (electron, photon, Higgs boson, ...) has its own quantum field.

#### 5.1.1 Spin 0: Scalar Field

E.g. Higgs boson, pions, ...

$$\phi(x) = \int \frac{d^3p}{\sqrt{(2\pi)^3 2E}} \left[ a(\vec{p}) e^{-ipx} + a^{c\dagger}(\vec{p}) e^{ipx} \right] . \quad (5.1)$$

### 5.1.2 Spin 1/2: the Dirac Field

E.g. quarks, leptons, ...

Note that single-particle Dirac equation is not exactly right even for single-particle systems (such as the Hydrogen atom) and unable to treat many-particle processes, such as the  $\beta$ -decay  $n \rightarrow p e^- \bar{\nu}$ . So we have to upgrade to quantum field theory.

Any Dirac field is some superposition of the complete set

$$u(\vec{p}, \sigma)e^{-ipx}, \quad v(\vec{p}, \sigma)e^{ipx}, \quad (\sigma = 1, 2, \quad \vec{p} \in \mathbb{R}^3) \quad (5.2)$$

and we can write it as:

$$\psi(x) = \sum_{\sigma} \int \frac{d^3p}{\sqrt{(2\pi)^3 2E}} [u(\vec{p}, \sigma)a(\vec{p}, \sigma)e^{-ipx} + v(\vec{p}, \sigma)a^{c\dagger}(\vec{p}, \sigma)e^{ipx}] . \quad (5.3)$$

Here  $1/\sqrt{(2\pi)^3 2E}$  is a normalization factor (there are many different conventions), and  $a(\vec{p}, \sigma)$  and  $a^{c\dagger}(\vec{p}, \sigma)$  are expansion coefficients. To make this a quantum Dirac field we promote these coefficients to the rank of operators by imposing the anticommutation relations

$$\{a(\vec{p}, \sigma), a^{\dagger}(\vec{p}', \sigma')\} = \delta_{\sigma\sigma'} \delta^3(\vec{p} - \vec{p}') , \quad (5.4)$$

and similarly for  $a^c(\vec{p}, \sigma)$ . For bosonic fields we would have a commutation relations instead. This is similar to the promotion of position and momentum to the rank of operators by the  $[x_i, p_j] = i\hbar\delta_{ij}$  commutation relations, which is why is this transition from the single-particle quantum theory to the quantum field theory sometimes called second quantization.

Operator  $a^{\dagger}$ , when operating on vacuum  $|0\rangle$ , creates one-particle state  $|\vec{p}, \sigma\rangle$ ,

$$a^{\dagger}(\vec{p}, \sigma)|0\rangle = |\vec{p}, \sigma\rangle , \quad (5.5)$$

and this is the reason that it is named a creation operator. Similarly,  $a$  is an annihilation operator

$$a(\vec{p}, \sigma)|\vec{p}, \sigma\rangle = |0\rangle , \quad (5.6)$$

and  $a^{c\dagger}$  and  $a^c$  are creation and annihilation operators for antiparticle states ( $c$  in  $a^c$  stands for 'conjugated').

The fact that Dirac spinors satisfy the Klein-Gordon equation suggests the *ansatz*

$$\psi(x) = u(\vec{p})e^{-ipx}, \quad (5.7)$$

which after inclusion in the Dirac equation gives the momentum space Dirac equation

$$(\not{p} - m)u(\vec{p}) = 0. \quad (5.8)$$

This has two positive-energy solutions

$$u(\vec{p}, \sigma) = N \begin{pmatrix} \chi^{(\sigma)} \\ \frac{\vec{\sigma} \cdot \vec{p}}{E + m} \chi^{(\sigma)} \end{pmatrix}, \quad (\sigma = 1, 2) \quad (5.9)$$

where

$$\chi^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (5.10)$$

and two negative-energy solutions which are then interpreted as positive-energy antiparticle solutions

$$v(\vec{p}, \sigma) = -N \begin{pmatrix} \frac{\vec{\sigma} \cdot \vec{p}}{E + m} (i\sigma^2)\chi^{(\sigma)} \\ (i\sigma^2)\chi^{(\sigma)} \end{pmatrix}. \quad (\sigma = 1, 2, \quad E > 0) \quad (5.11)$$

$N$  is the normalization constant to be determined later. The momentum-space Dirac equation for antiparticle solutions is

$$(\not{p} + m)v(\vec{p}, \sigma) = 0. \quad (5.12)$$

It can be shown that the two solutions, one with  $\sigma = 1$  and another with  $\sigma = 2$ , correspond to the two spin states of the spin-1/2 particle.

**Normalization:** In non-relativistic single-particle quantum mechanics normalization of a wavefunction is straightforward. Probability that the particle is somewhere in space is equal to one, and this translates into the normalization condition  $\int \psi^* \psi dV = 1$ . On the other hand, we will eventually use spinors (5.9) and (5.11) in many-particle quantum field theory so their normalization is not unique. We will choose normalization convention where we have  $N$  particles in the unit volume:

$$\int_{\text{unit volume}} \rho dV = \int_{\text{unit volume}} \psi^\dagger \psi dV = N. \quad (5.13)$$

This choice is relativistically covariant because the Lorentz contraction of the volume element is compensated by the energy change. There are other normalization conventions with other advantages.

**Parity and Bilinear Covariants:** The parity transformation:

$$P : (\vec{x} \rightarrow -\vec{x}, t \rightarrow t) , \quad \psi \rightarrow \gamma^0 \psi . \quad (5.14)$$

Any fermion current will be of the form  $\bar{\psi}\Gamma\psi$ , where  $\Gamma$  is some four-by-four matrix. For construction of interaction Lagrangian we want to use only those currents that have definite Lorentz transformation properties. We first define two new matrices:

$$\begin{aligned} \gamma^5 &\equiv i\gamma^0\gamma^1\gamma^2\gamma^3 \stackrel{\text{Dirac rep.}}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \{\gamma^5, \gamma^\mu\} = 0 , \\ \sigma^{\mu\nu} &\equiv \frac{i}{2}[\gamma^\mu, \gamma^\nu] , \quad \sigma^{\mu\nu} = -\sigma^{\nu\mu} . \end{aligned} \quad (5.15)$$

Now  $\bar{\psi}\Gamma\psi$  will transform covariantly if  $\Gamma$  is one of the matrices given in the following table. Transformation properties of  $\bar{\psi}\Gamma\psi$ , the number of different  $\gamma$  matrices in  $\Gamma$ , and the number of components of  $\Gamma$  are also displayed.

$\Gamma$	transforms as	number of $\gamma$ 's	number of components
1	scalar	0	1
$\gamma^\mu$	vector	1	4
$\sigma^{\mu\nu}$	tensor	2	6
$\gamma^5\gamma^\mu$	axial vector	3	4
$\gamma^5$	pseudoscalar	4	1

This exhausts all possibilities. The total number of components is 16, meaning that the set  $\{1, \gamma^\mu, \sigma^{\mu\nu}, \gamma^5\gamma^\mu, \gamma^5\}$  makes a complete basis for any four-by-four matrix. Such  $\bar{\psi}\Gamma\psi$  currents are called bilinear covariants.

### 5.1.3 Spin 1: Vector Field

Either for massive (e.g. W,Z weak bosons), or massless (e.g. photon),

$$\begin{aligned} A^\mu(x) &= \sum_\lambda \int \frac{d^3p}{\sqrt{(2\pi)^3 2E}} [\epsilon^\mu(\vec{p}, \lambda) a(\vec{p}, \lambda) e^{-ipx} + \\ &+ \epsilon^{\mu*}(\vec{p}, \lambda) a^\dagger(\vec{p}, \lambda) e^{ipx}] , \end{aligned} \quad (5.16)$$

where  $\epsilon^\mu(\vec{p}, \lambda)$  is a polarization vector. For massive particles it obeys

$$p_\mu \epsilon^\mu(\vec{p}, \lambda) = 0 \quad (5.17)$$

automatically, whereas in the massless case this condition can be imposed thanks to gauge invariance (Lorentz gauge condition). This means that there are only three independent polarizations of a massive vector particle:  $\lambda = 1, 2, 3$  (or  $\lambda = +, -, 0$ ). In massless case gauge symmetry can be further exploited to eliminate one more polarization state leaving us with only two:  $\lambda = 1, 2$  (or  $\lambda = +, -$ ).

Normalization of polarization vectors is such that

$$\epsilon^*(\vec{p}, \lambda) \cdot \epsilon(\vec{p}, \lambda) = -1 . \quad (5.18)$$

For a massive particle moving along the  $z$ -axis ( $p = (E, 0, 0, |\vec{p}|)$ ) we can take

$$\epsilon(\vec{p}, \pm) = \mp \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ \pm i \\ 0 \end{pmatrix}, \quad \epsilon(\vec{p}, 0) = \frac{1}{m} \begin{pmatrix} |\vec{p}| \\ 0 \\ 0 \\ E \end{pmatrix}. \quad (5.19)$$

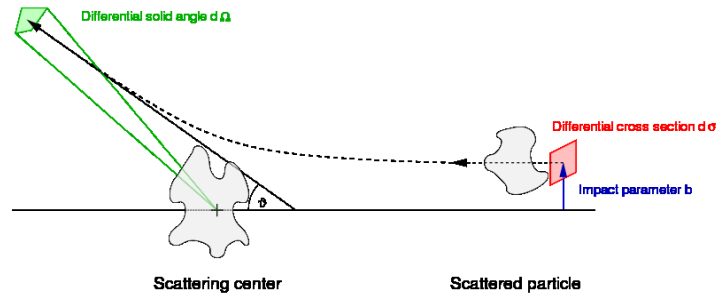
The obtained result obviously cannot be simply extrapolated to the massless case via the limit  $m \rightarrow 0$ . Gauge symmetry makes massless polarization sum somewhat more complicated but for the purpose of the simple Feynman diagram calculations it is permissible to use just the following relation

$$\sum_{\lambda} \epsilon^{\mu*}(\vec{p}, \lambda) \epsilon^{\nu}(\vec{p}, \lambda) = -g^{\mu\nu} . \quad (5.20)$$

## 5.2 Differential Cross Section

Consider a classical measurement where a single particle is scattered off a single stationary target particle. Conventionally, a spherical coordinate system is used, with the target placed at the origin and the  $z$  axis of this coordinate system aligned with the incident beam. The angle  $\theta$  is the scattering angle, measured between the incident beam and the scattered beam, and the  $\varphi$  is the azimuthal angle. The impact parameter  $b$  is the perpendicular offset of the trajectory of the incoming particle, and the outgoing particle emerges at an angle  $\theta$ . For a given interaction (Coulombic, magnetic, gravitational,

contact, etc.), the impact parameter and the scattering angle have a definite one-to-one functional dependence on each other. Generally the impact parameter can neither be controlled nor measured from event to event and is assumed to take all possible values when averaging over many scattering events.



The differential size of the cross section is the area element in the plane of the impact parameter,

$$d\sigma = b d\varphi db . \quad (5.21)$$

The differential angular range of the scattered particle at angle  $\theta$  is the solid angle element

$$d\Omega = \sin\theta d\theta d\varphi . \quad (5.22)$$

The differential cross section is the quotient of these quantities,  $d\sigma/d\Omega$ . It is a function of the scattering angle (and therefore also the impact parameter), plus other observables such as the momentum of the incoming particle. The differential cross section is always taken to be positive, even though larger impact parameters generally produce less deflection. In cylindrically symmetric situations (about the beam axis), the azimuthal angle  $\varphi$  is not changed by the scattering process, and  $d\sigma/d\Omega$  can be written as

$$\frac{d\sigma}{d\cos\theta} = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\sigma}{d\Omega} d\varphi . \quad (5.23)$$

In situations where the scattering process is not azimuthally symmetric, such as when the beam or target particles possess magnetic moments oriented perpendicular to the beam axis, the differential cross section must also be expressed as a function of the azimuthal angle.

For scattering of particles of incident flux  $F_{\text{inc}}$  off a stationary target consisting of many particles, the differential cross section  $d\sigma/d\Omega$  at an angle  $(\theta, \varphi)$

is related to the flux of scattered particle detection  $F_{\text{out}}(\theta, \varphi)$  in particles per unit time by

$$\frac{d\sigma}{d\Omega}(\theta, \varphi) = \frac{1}{nt\Delta\Omega} \frac{F_{\text{out}}(\theta, \varphi)}{F_{\text{inc}}} . \quad (5.24)$$

Here  $\Delta\Omega$  is the finite angular size of the detector,  $n$  is the number density of the target particles, and  $t$  is the thickness of the stationary target. This formula assumes that the target is thin enough that each beam particle will interact with at most one target particle.

The total cross section  $\sigma$  may be recovered by integrating the differential cross section  $d\sigma/d\Omega$  over the full solid angle ( $4\pi$  steradians):

$$\sigma = \oint_{4\pi} \frac{d\sigma}{d\Omega} d\Omega = \int_0^{2\pi} \int_0^\pi \frac{d\sigma}{d\Omega} \sin\theta d\theta d\varphi. \quad (5.25)$$

It is common to omit the "differential" qualifier when the type of cross section can be inferred from context. In this case,  $\sigma$  may be referred to as the integral cross section or total cross section. The latter term may be confusing in contexts where multiple events are involved, since "total" can also refer to the sum of cross sections over all events.

The differential cross section is extremely useful quantity in many fields of physics, as measuring it can reveal a great amount of information about the internal structure of the target particles. For example, the differential cross section of Rutherford scattering provided strong evidence for the existence of the atomic nucleus. Instead of the solid angle, the momentum transfer may be used as the independent variable of differential cross sections. Differential cross sections in inelastic scattering contain resonance peaks that indicate the creation of metastable states and contain information about their energy and lifetime.

### 5.3 Fermi Golden Rules

Principal experimental observables of particle physics are: the scattering cross section  $\sigma(1 + 2 \rightarrow 1' + 2' + \dots + n')$  and the decay width  $\Gamma(1 \rightarrow 1' + 2' + \dots + n')$ . On the other hand, theory is defined in terms of Lagrangian density of quantum fields, e.g.

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 - \frac{g}{4!} \phi^4 .$$

How to calculate  $\sigma$ 's and  $\Gamma$ 's from  $\mathcal{L}$ ?

To calculate rate of transition from the state  $|\alpha\rangle$  to the state  $|\beta\rangle$  in the presence of the interaction potential  $V_I$  in non-relativistic quantum theory we have *the Fermi's Golden Rule*:

$$\left( \begin{array}{c} \alpha \rightarrow \beta \\ \text{transition rate} \end{array} \right) = \frac{2\pi}{\hbar} |\langle \beta | V_I | \alpha \rangle|^2 \times \left( \begin{array}{c} \text{density of final} \\ \text{quantum states} \end{array} \right). \quad (5.26)$$

This is in the lowest order perturbation theory. For higher orders we have terms with products of more interaction potential matrix elements  $\langle |V_I| \rangle$ .

In quantum field theory there is a counterpart to these matrix elements – *the S-matrix*:

$$\langle \beta | V_I | \alpha \rangle + (\text{higher-order terms}) \longrightarrow \langle \beta | S | \alpha \rangle. \quad (5.27)$$

On one side,  $S$ -matrix elements can be perturbatively calculated (knowing the interaction Lagrangian/Hamiltonian) with the help of *the Dyson series*:

$$S = 1 - i \int d^4x_1 \mathcal{H}(x_1) + \frac{(-i)^2}{2!} \int d^4x_1 d^4x_2 T \{ \mathcal{H}(x_1) \mathcal{H}(x_2) \} + \dots, \quad (5.28)$$

and on another, we have 'golden rules' that associate these matrix elements with cross-sections and decay widths.

It is convenient to express these golden rules in terms of *the Feynman invariant amplitude*  $\mathcal{M}$  which is obtained by stripping some kinematical factors off the  $S$ -matrix:

$$\langle \beta | S | \alpha \rangle = \delta_{\beta\alpha} - i(2\pi)^4 \delta^4(p_\beta - p_\alpha) \mathcal{M}_{\beta\alpha} \prod_{i=\alpha,\beta} \frac{1}{\sqrt{(2\pi)^3 2E_i}}. \quad (5.29)$$

Now we have two rules:

- Partial decay rate of  $1 \rightarrow 1' + 2' + \dots + n'$ ,

$$d\Gamma = \frac{1}{2E_1} |\overline{\mathcal{M}_{\beta\alpha}}|^2 (2\pi)^4 \delta^4(p_1 - p'_1 - \dots - p'_n) \prod_{i=1}^n \frac{d^3p'_i}{(2\pi)^3 2E'_i}; \quad (5.30)$$

- Differential cross section for a scattering  $1 + 2 \rightarrow 1' + 2' + \dots + n'$ ,

$$\begin{aligned} d\sigma &= \frac{1}{u_\alpha} \frac{1}{2E_1} \frac{1}{2E_2} |\overline{\mathcal{M}_{\beta\alpha}}|^2 (2\pi)^4 \times \\ &\times \delta^4(p_1 + p_2 - p'_1 - \dots - p'_n) \prod_{i=1}^n \frac{d^3p'_i}{(2\pi)^3 2E'_i}, \quad (5.31) \end{aligned}$$



where  $u_\alpha$  is the relative velocity of particles 1 and 2:

$$u_\alpha = \frac{\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}}{E_1 E_2}, \quad (5.32)$$

and  $|\overline{\mathcal{M}}|^2$  is the Feynman invariant amplitude averaged over unmeasured particle spins (see Section 5.5.1). The dimension of  $\mathcal{M}$ , in units of energy, is for decays  $[\mathcal{M}] = 3 - n$ , and for scattering of two particles  $[\mathcal{M}] = 2 - n$ , where  $n$  is the number of produced particles.

For  $2 \rightarrow 2$  scattering we have in the center of the mass:

$$d\sigma_{\text{CM}}(1 + 2 \rightarrow 3 + 4) = \frac{1}{64\pi^2(E_1 + E_2)^2} \frac{|\vec{p}_3|}{|\vec{p}_1|} |\overline{\mathcal{M}_{\beta\alpha}}|^2 d\Omega_3, \quad (5.33)$$

where  $\Omega_3$  is the spatial angle of the third particle with respect to the first one ( $d\Omega_3 = d\phi_3 d\cos\theta_3$ ). So calculation of some observable quantity consists of two stages:

1. Determination of  $|\overline{\mathcal{M}}|^2$ . For this we use the method of Feynman diagrams to be introduced in the next section;
2. Integration over the Lorentz invariant phase space

$$d\Theta = (2\pi)^4 \delta^4(p_1 + p_2 - p'_1 - \dots - p'_n) \prod_{i=1}^n \frac{d^3 p'_i}{(2\pi)^3 2E'_i}.$$

## 5.4 Feynman Diagrams for $\phi^4$ -model

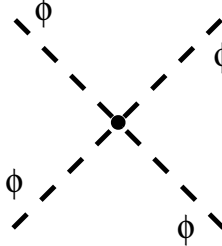
Consider the example of calculations of Feynman diagrams for the  $\phi^4$ -theory lagrangian:

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 - \frac{g}{4!} \phi^4. \quad (5.34)$$

- Free (kinetic) Lagrangian (terms with exactly two fields) determines particles of the theory and their propagators. Here we have just one scalar field:



- Interaction Lagrangian (terms with three or more fields) determines possible vertices. Here, again, there is just one vertex:



We need to construct all possible diagrams with fixed outer particles. For example, for scattering of two scalar particles in this theory we would have:

$$\mathcal{M}(1 + 2 \rightarrow 3 + 4) = \text{[tree diagram]} + \text{[loop diagram]} + \text{[loop diagram]} + \dots$$

In these diagrams time flows from left to right. Some people draw Feynman diagrams with time flowing up, which is more in accordance with the way we usually draw space-time in relativity physics.

Since each vertex corresponds to one interaction Lagrangian (Hamiltonian) term in (5.28), diagrams with loops correspond to higher orders of perturbation theory. Here we will work only to the lowest order, so we will use *tree diagrams* only.

To actually write down the Feynman amplitude  $\mathcal{M}$ , we have a set of Feynman rules that associate factors with elements of the Feynman diagram. In particular, to get  $-i\mathcal{M}$  we construct the Feynman rules in the following way:

- the vertex factor is just the  $i$  times the interaction term in the (momentum space) Lagrangian with all fields removed:

$$i\mathcal{L}_1 = -i\frac{g}{4!}\phi^4 \xrightarrow{\text{removing fields}} \text{[vertex diagram]} = -i\frac{g}{4!}; \quad (5.35)$$

- the propagator is  $i$  times the inverse of the kinetic operator (defined by the free equation of motion) in the momentum space:

$$\mathcal{L}_{\text{free}} \xrightarrow{\text{Euler eq.}} (\partial_\mu \partial^\mu + m^2)\phi = 0. \quad (5.36)$$

Going to the momentum space using the substitution  $\partial^\mu \rightarrow -ip^\mu$  and then taking the inverse gives:

$$(p^2 - m^2)\phi = 0 \quad \rightarrow \quad \bullet \text{---} \overset{\phi}{\text{---}} \text{---} \bullet = \frac{i}{p^2 - m^2} \quad (5.37)$$

(actually, the correct Feynman propagator is  $i/(p^2 - m^2 + i\epsilon)$ , but for our purposes we can ignore the infinitesimal  $i\epsilon$  term);

- External lines are represented by the appropriate polarization vector or spinor (the one that stands by the appropriate creation or annihilation operator in the fields (5.1), (5.3), (5.16) and their conjugates):

particle	Feynman rule
ingoing fermion	$u$
outgoing fermion	$\bar{u}$
ingoing antifermion	$\bar{v}$
outgoing antifermion	$v$
ingoing photon	$\epsilon^\mu$
outgoing photon	$\epsilon^{\mu*}$
ingoing scalar	1
outgoing scalar	1

The tree-level contribution to the scalar-scalar scattering amplitude in this  $\phi^4$  theory would be just

$$-i\mathcal{M} = -i\frac{g}{4!}. \quad (5.38)$$

Note that also

$$\bullet \xrightarrow{\mathbf{p}} \bullet = \frac{i \sum_\sigma u(\vec{p}, \sigma) \bar{u}(\vec{p}, \sigma)}{p^2 - m^2}, \quad (5.39)$$

i.e. the electron propagator is just the scalar propagator multiplied by the polarization sum. It is nice that this generalizes to propagators of all particles. This is very helpful since inverting the photon kinetic operator is

non-trivial due to gauge symmetry complications. Hence, propagators of vector particles are

$$\begin{aligned}
 \text{massive:} & \quad \text{---} \overset{\mathbf{p}, m}{\text{wavy}} \text{---} = \frac{-i \left( g^{\mu\nu} - \frac{p^\mu p^\nu}{m^2} \right)}{p^2 - m^2}, \\
 \text{massless:} & \quad \text{---} \overset{\mathbf{p}}{\text{wavy}} \text{---} = \frac{-i g^{\mu\nu}}{p^2}. \tag{5.40}
 \end{aligned}$$

This is in principle almost all we need to know to be able to calculate the Feynman amplitude of any given process. Note that propagators and external-line polarization vectors are determined only by the particle type (its spin and mass) so that the corresponding rules above are not restricted only to the  $\phi^4$  theory and QED, but will apply to all theories of scalars, spin-1 vector bosons and Dirac fermions (such as the standard model). The only additional information we need are the vertex factors.

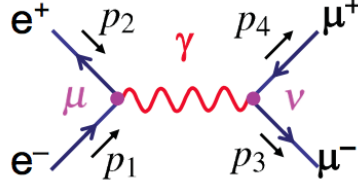
Almost in the preceding paragraph alludes to the fact that in general Feynman diagram calculation there are several additional subtleties:

- In loop diagrams some internal momenta are undetermined and we have to integrate over those. Also, there is an additional factor (-1) for each closed fermion loop. Since we will consider tree-level diagrams only, we can ignore this;
- There are some combinatoric numerical factors when identical fields come into a single vertex;
- Sometimes there is a relative (-) sign between diagrams;
- There is a symmetry factor if there are identical particles in the final state.

Note that there are several computer programs that can perform some or all of the steps in the calculation of Feynman diagrams. One such program is `FeynCalc` package for Wolfram's *Mathematica*.

## 5.5 Calculating $e^+e^- \rightarrow \mu^+\mu^-$ in QED

There is only one contributing tree-level diagram:



We write down the amplitude using the Feynman rules of QED and following fermion lines backwards. Order of lines themselves is unimportant.

$$i\mathcal{M} = [\bar{u}(\vec{p}_3, \sigma_3)(ie\gamma^\nu)v(\vec{p}_4, \sigma_4)] \frac{-ig_{\mu\nu}}{(p_1 + p_2)^2} \times [\bar{v}(\vec{p}_2, \sigma_2)(ie\gamma^\mu)u(\vec{p}_1, \sigma_1)] , \quad (5.41)$$

or, introducing abbreviation  $u_1 \equiv u(\vec{p}_1, \sigma_1)$ ,

$$\mathcal{M} = \frac{e^2}{(p_1 + p_2)^2} [\bar{u}_3\gamma_\mu v_4][\bar{v}_2\gamma^\mu u_1] . \quad (5.42)$$

### 5.5.1 Summing Over Polarizations

If we knew momenta and polarizations of all external particles, we could calculate  $\mathcal{M}$  explicitly. However, experiments are often done with unpolarized particles so we have to sum over the polarizations (spins) of the final particles and average over the polarizations (spins) of the initial ones:

$$|\mathcal{M}|^2 \rightarrow \overline{|\mathcal{M}|^2} = \underbrace{\frac{1}{2} \frac{1}{2} \sum_{\sigma_1 \sigma_2}}_{\text{avg. over initial pol.}} \overbrace{\sum_{\sigma_3 \sigma_4}}^{\text{sum over final pol.}} |\mathcal{M}|^2 . \quad (5.43)$$

Factors 1/2 are due to the fact that each initial fermion has two polarization (spin) states.

In the calculation of  $|\mathcal{M}|^2 = \mathcal{M}^* \mathcal{M}$ , the following identity is needed

$$[\bar{u}\gamma^\mu v]^* = [u^\dagger \gamma^0 \gamma^\mu v]^\dagger = v^\dagger \gamma^{\mu\dagger} \gamma^0 u = [\bar{v}\gamma^\mu u] . \quad (5.44)$$

Thus,

$$\overline{|\mathcal{M}|^2} = \frac{e^4}{4(p_1 + p_2)^4} \sum_{\sigma_{1,2,3,4}} [\bar{v}_4\gamma_\mu u_3][\bar{u}_1\gamma^\mu v_2][\bar{u}_3\gamma_\nu v_4][\bar{v}_2\gamma^\nu u_1] . \quad (5.45)$$

### 5.5.2 Casimir Trick

Sums over polarizations are easily performed using the following trick. First we write  $\sum[\bar{u}_1\gamma^\mu v_2][\bar{v}_2\gamma^\nu u_1]$  with explicit spinor indices:

$$\sum_{\sigma_1\sigma_2} \bar{u}_{1\alpha}\gamma^\mu_{\alpha\beta}v_{2\beta} \bar{v}_{2\gamma}\gamma^\nu_{\gamma\delta}u_{1\delta} . \quad (\alpha, \beta, \gamma, \delta = 1, 2, 3, 4) \quad (5.46)$$

We can now move  $u_{1\delta}$  to the front ( $u_{1\delta}$  is just a number, element of  $u_1$  vector, so it commutes with everything), and then use the completeness relations,

$$\begin{aligned} \sum_{\sigma_1} u_{1\delta} \bar{u}_{1\alpha} &= (\not{p}_1 + m_1)_{\delta\alpha} , \\ \sum_{\sigma_2} v_{2\beta} \bar{v}_{2\gamma} &= (\not{p}_2 - m_2)_{\beta\gamma} , \end{aligned} \quad (5.47)$$

which turn sum (5.46) into

$$(\not{p}_1 + m_1)_{\delta\alpha} \gamma^\mu_{\alpha\beta} (\not{p}_2 - m_2)_{\beta\gamma} \gamma^\nu_{\gamma\delta} = \text{Tr}[(\not{p}_1 + m_1)\gamma^\mu(\not{p}_2 - m_2)\gamma^\nu] . \quad (5.48)$$

This means that

$$\begin{aligned} \overline{|\mathcal{M}|^2} &= \frac{e^4}{4(p_1 + p_2)^4} \text{Tr}[(\not{p}_1 + m_1)\gamma^\mu(\not{p}_2 - m_2)\gamma^\nu] \times \\ &\times \text{Tr}[(\not{p}_4 - m_4)\gamma_\mu(\not{p}_3 + m_3)\gamma_\nu] . \end{aligned} \quad (5.49)$$

Thus we got rid off all the spinors and we are left only with traces of  $\gamma$  matrices. These can be evaluated using the relations from the following section.

### 5.5.3 Traces and Contraction Identities of $\gamma$ -Matrices

Traces and Contraction Identities are consequence of the anticommutation relations  $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ ,  $\{\gamma^\mu, \gamma^5\} = 0$ ,  $(\gamma^5)^2 = 1$ .

**Trace Identities:**

- Trace of an odd number of  $\gamma$ 's vanishes:

$$\begin{aligned}
\text{Tr}(\gamma^{\mu_1}\gamma^{\mu_2}\dots\gamma^{\mu_{2n+1}}) &= \text{Tr}(\gamma^{\mu_1}\gamma^{\mu_2}\dots\gamma^{\mu_{2n+1}}\overbrace{\gamma^5\gamma^5}^1) \\
\text{(moving } \gamma^5 \text{ over each } \gamma^{\mu_i}) &= -\text{Tr}(\gamma^5\gamma^{\mu_1}\gamma^{\mu_2}\dots\gamma^{\mu_{2n+1}}\gamma^5) \\
\text{(cyclic property of trace)} &= -\text{Tr}(\gamma^{\mu_1}\gamma^{\mu_2}\dots\gamma^{\mu_{2n+1}}\gamma^5\gamma^5) \\
&= -\text{Tr}(\gamma^{\mu_1}\gamma^{\mu_2}\dots\gamma^{\mu_{2n+1}}) \\
&= 0;
\end{aligned}$$

- $\text{Tr } 1 = 4$ ;

- We have:

$$\begin{aligned}
\text{Tr}\gamma^\mu\gamma^\nu &= \text{Tr}(2g^{\mu\nu} - \gamma^\nu\gamma^\mu) \stackrel{(2.)}{=} 8g^{\mu\nu} - \text{Tr}\gamma^\nu\gamma^\mu = \\
&= 8g^{\mu\nu} - \text{Tr}\gamma^\mu\gamma^\nu \rightarrow 2\text{Tr}\gamma^\mu\gamma^\nu = 8g^{\mu\nu} \rightarrow \text{Tr}\gamma^\mu\gamma^\nu = 4g^{\mu\nu}. \quad (5.50)
\end{aligned}$$

This also implies:

$$\text{Tr}\not{a}\not{b} = 4a \cdot b. \quad (5.51)$$

### Contraction Identities:

- 

$$\gamma^\mu\gamma_\mu = \frac{1}{2}g_{\mu\nu}\underbrace{(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu)}_{2g^{\mu\nu}} = g_{\mu\nu}g^{\mu\nu} = 4;$$

- 

$$\underbrace{\gamma^\mu\gamma^\alpha\gamma_\mu}_{-\gamma_\mu\gamma^\alpha+2g_\mu^\alpha} = -4\gamma^\alpha + 2\gamma^\alpha = -2\gamma^\alpha.$$

### 5.5.4 Kinematics in the CM Frame

In  $e^+e^-$  colliders often  $p_i \gg m_e, m_\mu$ ,  $i = 1, \dots, 4$ , so we can take

$$m_i \rightarrow 0 \quad \text{'high-energy' or 'extreme relativistic' limit.}$$

Then

$$\overline{|\mathcal{M}|^2} = \frac{8e^4}{(p_1 + p_2)^4} [(p_1 \cdot p_3)(p_2 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3)]. \quad (5.52)$$

### 5.5.5 Integration Over 2-Particle Phase Space

Now we can use the 'golden rule' (5.31) for the  $1 + 2 \rightarrow 3 + 4$  differential scattering cross-section

$$d\sigma = \frac{1}{u_\alpha} \frac{1}{2E_1} \frac{1}{2E_2} |\overline{\mathcal{M}}|^2 d\Theta_2, \quad (5.53)$$

where two-particle phase space to be integrated over is

$$d\Theta_2 = (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \frac{d^3 p_3}{(2\pi)^3 2E_3} \frac{d^3 p_4}{(2\pi)^3 2E_4}. \quad (5.54)$$

First we integrate over four out of six integration variables, and we do this in general frame.  $\delta$ -function makes the integration over  $d^3 p_4$  trivial giving

$$d\Theta_2 = \frac{1}{(2\pi)^2 4E_3 E_4} \delta(E_1 + E_2 - E_3 - E_4) \underbrace{d^3 p_3}_{\vec{p}_3^2 d|\vec{p}_3| d\Omega_3}. \quad (5.55)$$

Now we integrate over  $d|\vec{p}_3|$  by noting that  $E_3$  and  $E_4$  are functions of  $|\vec{p}_3|$

$$E_3 = E_3(|\vec{p}_3|) = \sqrt{\vec{p}_3^2 + m_3^2}, \quad E_4 = \sqrt{\vec{p}_4^2 + m_4^2} = \sqrt{\vec{p}_3^2 + m_4^2}, \quad (5.56)$$

and by  $\delta$ -function relation

$$\begin{aligned} \delta(E_1 + E_2 - \sqrt{\vec{p}_3^2 + m_3^2} - \sqrt{\vec{p}_3^2 + m_4^2}) &= \\ &= \delta[f(|\vec{p}_3|)] = \frac{\delta(|\vec{p}_3| - |\vec{p}_3^{(0)}|)}{|f'(|\vec{p}_3|)|_{|\vec{p}_3|=|\vec{p}_3^{(0)}|}}. \end{aligned} \quad (5.57)$$

Here  $|\vec{p}_3|$  is just the integration variable and  $|\vec{p}_3^{(0)}|$  is the zero of  $f(|\vec{p}_3|)$  i.e. the actual momentum of the third particle. After we integrate over  $d|\vec{p}_3|$  we put  $|\vec{p}_3^{(0)}| \rightarrow |\vec{p}_3|$ . Since

$$f'(|\vec{p}_3|) = -\frac{E_3 + E_4}{E_3 E_4} |\vec{p}_3|, \quad (5.58)$$

we get

$$d\text{Lips}_2 = \frac{|\vec{p}_3| d\Omega}{16\pi^2 (E_1 + E_2)}. \quad (5.59)$$

Now we again specialize to the CM frame and note that the flux factor is

$$4E_1 E_2 u_\alpha = 4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = 4|\vec{p}_1|(E_1 + E_2), \quad (5.60)$$



giving finally

$$\frac{d\sigma_{\text{CM}}}{d\Omega} = \frac{1}{64\pi^2(E_1 + E_2)^2} \frac{|\vec{p}_3|}{|\vec{p}_1|} |\mathcal{M}|^2. \quad (5.61)$$

Note that we kept masses in each step so this formula is generally valid for any CM scattering. For our particular  $e^-e^+ \rightarrow \mu^-\mu^+$  scattering this gives the final result for differential cross-section (introducing the fine structure constant  $\alpha = e^2/(4\pi)$ )

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{16E^2} (1 + \cos^2 \theta). \quad (5.62)$$

Note that it is obvious that  $\sigma \propto \alpha^2$ , and that dimensional analysis requires  $\sigma \propto 1/E^2$ , so only angular dependence ( $1 + \cos^2 \theta$ ) tests QED as a theory of leptons and photons.

### 5.5.6 Summary of Steps

To recapitulate, calculating (unpolarized) scattering cross-section (or decay width) consists of the following steps:

1. drawing the Feynman diagram(s)
2. writing  $-i\mathcal{M}$  using the Feynman rules
3. squaring  $\mathcal{M}$  and using the Casimir trick to get traces
4. evaluating traces
5. applying kinematics of the chosen frame
6. integrating over the phase space

**Exercise 5.1:** Determine Dirac equations for  $\bar{u}(\vec{p}, \sigma)$  and  $\bar{v}(\vec{p}, \sigma)$ .

**Exercise 5.2:** Using the explicit expressions (5.9) and (5.11) show that

$$\sum_{\sigma=1,2} u(\vec{p}, \sigma)\bar{u}(\vec{p}, \sigma) = \not{p} + m, \quad \sum_{\sigma=1,2} v(\vec{p}, \sigma)\bar{v}(\vec{p}, \sigma) = \not{p} - m.$$

**Exercise 5.3:** For a vector field calculate  $\sum_{\lambda} \epsilon^{\mu*}(\vec{p}, \lambda)\epsilon^{\nu}(\vec{p}, \lambda)$ . Hint: Write it in the most general form  $(Ag^{\mu\nu} + Bp^{\mu}p^{\nu})$  and then determine  $A$  and  $B$ .

**Exercise 5.4:** Determine the Feynman rules for the electron propagator and for the only vertex of quantum electrodynamics.

**Exercise 5.5:** Express  $|\overline{\mathcal{M}}|^2$  for  $e^-e^+ \rightarrow \mu^-\mu^+$  scattering in terms of Mandelstam variables.

**Exercise 5.6:** Draw Feynman diagram(s) and write down the amplitude for Compton scattering  $\gamma e^- \rightarrow \gamma e^-$ .

**Exercise 5.7:** Why we sum probabilities and not amplitudes?

**Exercise 5.8:** Calculate  $\text{Tr}(\gamma^\mu \gamma^\nu \gamma^\rho \gamma^\sigma)$ . Hint: Move  $\gamma^\sigma$  all the way to the left, using the anticommutation relations.

**Exercise 5.9:**

1. Prove that  $\text{Tr}(\gamma^{\mu_1} \gamma^{\mu_2} \dots \gamma^{\mu_{2n}})$  has  $(2n - 1)!!$  terms.
2.  $\text{Tr}(\gamma^5 \gamma^{\mu_1} \gamma^{\mu_2} \dots \gamma^{\mu_{2n+1}}) = 0$ . This follows from 1. and from the fact that  $\gamma^5$  consists of even number of  $\gamma$ 's.
3.  $\text{Tr} \gamma^5 = \text{Tr}(\gamma^0 \gamma^0 \gamma^5) = -\text{Tr}(\gamma^0 \gamma^5 \gamma^0) = -\text{Tr} \gamma^5 = 0$ .
4.  $\text{Tr}(\gamma^5 \gamma^\mu \gamma^\nu) = 0$ . (Same trick as above, with  $\gamma^\alpha \neq \mu, \nu$  instead of  $\gamma^0$ .)
5.  $\text{Tr}(\gamma^5 \gamma^\mu \gamma^\nu \gamma^\rho \gamma^\sigma) = -4i \epsilon^{\mu\nu\rho\sigma}$ , with  $\epsilon^{0123} = 1$ . Careful: convention with  $\epsilon^{0123} = -1$  is also in use.

**Exercise 5.10:**

1. Contract  $\gamma^\mu \gamma^\alpha \gamma^\beta \gamma_\mu$ .
2. Show that  $\gamma^\mu \gamma^\alpha \gamma^\beta \gamma^\gamma \gamma_\mu = -2\gamma^\gamma \gamma^\beta \gamma^\alpha$ .
3. For  $e^+e^- \rightarrow \mu^+\mu^-$  calculate traces in  $|\overline{\mathcal{M}}|^2$ :  $\text{Tr}[(\not{p}_1 + m_1)\gamma^\mu(\not{p}_2 - m_2)\gamma^\nu]$  and  $\text{Tr}[(\not{p}_4 - m_4)\gamma_\mu(\not{p}_3 + m_3)\gamma_\nu]$ .
4. Calculate  $|\overline{\mathcal{M}}|^2$  for  $e^+e^- \rightarrow \mu^+\mu^-$ .

**Exercise 5.11:** Express  $|\overline{\mathcal{M}}|^2$  in terms of  $E$  and  $\theta$ .

**Exercise 5.12:** Integrate (5.62) to get the total cross section  $\sigma$ .

## Chapter 6

# Conservation Laws and Selection Rules

In developing the standard model for particles, certain types of interactions and decays are observed to be common and others seem to be forbidden. The study of interactions has led to a number of conservation laws which govern them. These conservation laws are in addition to the classical conservation laws such as conservation of energy, charge, etc., which still apply in the realm of particle interactions. Strong overall conservation laws are the conservation of baryon and lepton numbers. Specific quantum numbers have been assigned to the different fundamental particles, and other conservation laws are associated with those quantum numbers.

### 6.1 Crossing Symmetry

Part of the high energy physicist's toolkit in anticipating what interactions can be expected is "crossing symmetry". Any interaction which is observed can be used to predict other related interactions by "crossing" any particle across the reaction symbol and turning it into its antiparticle.

If a particle interaction

$$A + B \rightarrow C + D \tag{6.1}$$

is observed to occur, then related interactions can be anticipated from the

fact that any of the particles can be replaced by its antiparticle on the other side of the interaction. This is commonly known as "crossing symmetry". The observation of the above interaction implies the existence of the following interactions.

$$\begin{aligned}
 A &\rightarrow \bar{B} + C + D , \\
 A + \bar{C} &\rightarrow \bar{B} + D , \\
 \bar{C} &\rightarrow \bar{A} + \bar{B} + D , \\
 \bar{C} + \bar{D} &\rightarrow \bar{A} + \bar{B} .
 \end{aligned}
 \tag{6.2}$$

The overbar indicates the antiparticle. Crossing symmetry applies to all known particles, including the photon which is its own antiparticle. One example of the crossing principle is that of the relation between Compton scattering and electron-positron annihilation.

$$\begin{aligned}
 \text{Compton scattering :} &\quad \gamma + e^- \rightarrow e^- + \gamma , \\
 \text{Pair annihilation :} &\quad e^- + e^+ \rightarrow \gamma + \gamma .
 \end{aligned}
 \tag{6.3}$$

By examination, it can be seen that these two interactions are related by crossing symmetry. It could then be said that the observation of Compton scattering implies the existence of pair annihilation and predicts that it will produce a pair of photons.

Another example of crossing symmetry may have led Reines and Cowan to their experiment for the detection of the neutrino. If you take the electron product from the neutron decay reaction to the other side and convert it into a positron, then you have the reaction which they used,

$$\begin{aligned}
 \text{Neutron decay :} &\quad n \rightarrow p + e^- + \bar{\nu}_e , \\
 \text{Neutrino detection :} &\quad \bar{\nu}_e + p \rightarrow n + e^+ .
 \end{aligned}
 \tag{6.4}$$

## 6.2 Totalitarian Principle

From what we observe with massive particles, it would seem that any localized particle of finite mass should be unstable, since the decay into several smaller particles provides many more ways to distribute the energy, and thus would have higher entropy. Some have adopted the description "totalitarian principle" for this situation. It might be stated as "every process that is not

forbidden must occur". From this point of view, any decay process which is expected but not observed must be prevented from occurring by some conservation law. This approach has been fruitful in helping to determine the rules for particle decay.

For example, with just conservation of energy and charge considered, one might expect a proton to decay into a positive pion plus a gamma ray to take away excess energy and conserve momentum:

$$p \rightarrow \pi^+ + \gamma . \quad (6.5)$$

The fact that neither this nor any other decay of the proton is observed suggests that the decay of the proton is forbidden by a strong conservation principle. This principle is called the conservation of baryon number, and no observed particle decays violate it. The proton does not decay because it is the least massive baryon, and has nowhere to go.

Another decay which was expected on energy and charge grounds was the decay of the neutron into a proton and an electron. The decay of the neutron is observed, but the fact that the electron does not have a definite energy implies that there is a third particle in the decay, the antineutrino,

$$\begin{aligned} n &\rightarrow p + e^- , \\ n &\rightarrow p + e^- + \bar{\nu}_e . \end{aligned} \quad (6.6)$$

The fact that the first of these decays did not occur suggested a prohibiting conservation law, which is called the conservation of lepton number.

Since the strengths of the interactions associated with particle decay descend in the order strong, electromagnetic and weak, it might be presumed that the strongest interaction would lead to the shortest lifetime, and that is what is observed. From experiments we can establish time regimes for the three types of interactions.

Interaction	Approximate decay lifetime
Strong	$10^{-23}$ s
Electromagnetic	$10^{-16}$ s
Weak	$10^{-10}$ s

In the spirit of the "totalitarian principle", if you observed a decay in the  $10^{-16}$  s range you might guess that it is electromagnetic, and that some

principle prevented it decay by the strong interaction. A  $10^{-10}$  s decay suggests that both strong and electromagnetic are somehow blocked.

One way to examine a decay is to list the quark content of each of the particles. For this purpose, it is convenient to refer to the meson table:

Mesons $q\bar{q}$					
Symbol	Name	Quark content	Electric charge	Mass $\text{GeV}/c^2$	Spin
$\pi^+$	pion	$u\bar{d}$	+1	0.140	0
$K^-$	kaon	$s\bar{u}$	-1	0.494	0
$\rho^+$	rho	$u\bar{d}$	+1	0.770	1
$B^0$	B-zero	$d\bar{b}$	0	5.279	0
$\eta_c$	eta-c	$c\bar{c}$	0	2.980	0

and baryon table:

Example Baryons ( $qqq$ ) and AntiBaryons ( $\bar{q}\bar{q}\bar{q}$ )					
Symbol	Name	Quark Content	Chg	Mass $\text{GeV}/c^2$	Spin
$p$	proton	$uud$	+e	0.938	1/2
$\bar{p}$	antiproton	$\bar{u}\bar{u}\bar{d}$	-e	0.938	1/2
$n$	neutron	$udd$	0	0.940	1/2
$\Lambda$	lambda	$uds$	0	1.116	1/2
$\Delta^{++}$	delta++	$uuu$	+2e	1.232	3/2
$\Omega^-$	omega-	$sss$	-e	1.672	3/2

### 6.3 Conservation of Baryon Number

Each of the baryons is assigned a baryon number  $B = 1$ . This can be considered to be equivalent to assigning each quark a baryon number of  $1/3$ . This implies that the mesons, with one quark and one antiquark, have a baryon number  $B = 0$ . No known decay process or interaction in nature changes the net baryon number.

The neutron and all heavier baryons decay directly to protons or eventually form protons, the proton being the least massive baryon. This implies that the proton has nowhere to go without violating the conservation of baryon number, so if the conservation of baryon number holds exactly, the proton is completely stable against decay. One prediction of grand unification of forces is that the proton would have the possibility of decay, so that possibility is being investigated experimentally.

Conservation of baryon number prohibits a decay of the type

$$\begin{aligned} p + n &\rightarrow p + \mu^+ + \mu^- , \\ B = 1 + 1 &\neq 1 + 0 + 0 , \end{aligned} \tag{6.7}$$

but with sufficient energy permits pair production in the reaction

$$\begin{aligned} p + n &\rightarrow p + n + \bar{p} + p , \\ B = 1 + 1 &\neq 1 + 1 - 1 + 1 . \end{aligned} \tag{6.8}$$

The fact that the decay

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{6.9}$$

is observed implied that there is no corresponding principle of conservation of meson number. The pion is a meson composed of a quark and an antiquark, and on the right side of the equation there are only leptons. (Equivalently, you could assign a baryon number of 0 to the meson.)

## 6.4 Conservation of Lepton Number

The conservation of lepton number ( $L$ ) is a little more complicated rule than the conservation of baryon number because there is a separate requirement for each of the three sets of leptons, the electron, muon and tau and their associated neutrinos.

The first significant example was found in the decay of the neutron. When the decay of the neutron into a proton and an electron was observed, it did not fit the pattern of two-particle decay. That is, the electron emitted does not have a definite energy as is required by conservation of energy and momentum for a two-body decay. This implied the emission of a third

particle, which we now identify as the electron antineutrino,

$$\begin{aligned} n &\rightarrow p + e^- , \\ n &\rightarrow p + e^- + \bar{\nu}_e . \end{aligned} \tag{6.10}$$

The assignment of a lepton number of  $L = 1$  to the electron and  $L = -1$  to the electron antineutrino keeps the lepton number equal to zero on both sides of the second reaction above, while the first reaction does not conserve lepton number.

The observation of the following two decay processes leads to the conclusion that there is a separate lepton number for muons which must also be conserved,

$$\begin{aligned} \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu , \\ \mu^- &\rightarrow e^- + \bar{\nu}_e + \nu_\mu . \end{aligned} \tag{6.11}$$

The first reaction above (decay of the pion) is known to be a two-body decay by the fact that a well-defined muon energy is observed from the decay. However, the decay of the muon into an electron produces a distribution of electron energies, showing that it is at least a three-body decay. In order for both electron lepton number and muon lepton number to be conserved, then the other particles must be an electron anti-neutrino and a muon neutrino.

## 6.5 Isospin

Isospin is a term introduced to describe groups of particles which have nearly the same mass, such as the proton and neutron. This doublet of particles is said to have isospin  $1/2$ , with projection  $+1/2$  for the proton and  $-1/2$  for the neutron. The three pions compose a triplet, suggesting isospin  $1$ . The projections are  $+1$  for the positive,  $0$  and  $-1$  for the neutral and negative pions. Isospin is used as an axis in particle diagrams, with strangeness being the other axis. Isospin is not really spin, and doesn't have the units of angular momentum - the spin term is tacked on because the addition of the isospins follows the same rules as spin.

Isospin is a dimensionless quantity associated with the fact that the strong interaction is independent of electric charge. Any two members of the proton-neutron isospin doublet experience the same strong interaction: proton-



proton, proton-neutron, neutron-neutron have the same strong force attraction.

At the quark level, the up and down quarks form an isospin doublet ( $I = 1/2$ ) and the projection  $I_3 = +1/2$  is assigned to the up quark and  $I_3 = -1/2$  to the down. (The subscript 3 is used here for the third component rather than the z used with spin and orbital angular momentum because most of the literature does so.) The other quarks are assigned isospin  $I = 0$ . Isospin is related to other quantum numbers for the particles by

$$\frac{q}{e} = I_3 + \frac{Y}{2}, \quad (6.12)$$

where  $q$  denotes charge ( $q/e$  is used to make it dimensionless),  $I_3$  is projection of the isotopic spin,  $Y = S + B$  is called hypercharge and  $S$  and  $B$  are strangeness and baryon number, respectively. This relationship is called the Gell Mann-Nishijima formula. Some references use  $T$  for isospin, but it appears that most use  $I$  for isospin and  $T$  for weak isospin.

Isospin is associated with a conservation law which requires strong interaction decays to conserve isospin, as illustrated by the process,

$$\begin{array}{l} uds \quad \quad uds \\ \Sigma^0 = \Lambda^0 + \gamma, \\ I = 1 \neq I = 0 \end{array} \quad (6.13)$$

which does not involve any transmutation of quarks, so would be expected to decay by strong interaction. However, it does not conserve isospin, and is observed to decay by the electromagnetic interaction, but not by the strong interaction. The experimental discrimination is made by the observation of its decay lifetime, presuming by the totalitarian principle that if it could decay by the strong interaction, it would.

### 6.5.1 Strangeness

In 1947 during a study of cosmic ray interactions, a product of a proton collision with a nucleus was found to live for a much longer time than expected:  $10^{-10}$  seconds instead of the expected  $10^{-23}$  seconds! This particle was named the lambda particle ( $\Lambda^0$ ) and the property which caused it to live so long was dubbed "strangeness" and that name stuck to be the name

of one of the quarks from which the lambda particle is constructed. The lambda is a baryon which is made up of three quarks: an up, a down and a strange quark.

The shorter lifetime of  $10^{-23}$  seconds was expected because the lambda as a baryon participates in the strong interaction, and that usually leads to such very short lifetimes. The long observed lifetime helped develop a new conservation law for such decays called the "conservation of strangeness". The presence of a strange quark in a particle is denoted by a quantum number  $S = -1$ . Particle decay by the strong or electromagnetic interactions preserve the strangeness quantum number. The decay process for the lambda particle must violate that rule, since there is no lighter particle which contains a strange quark - so the strange quark must be transformed to another quark in the process. That can only occur by the weak interaction, and that leads to a much longer lifetime. The decay processes show that strangeness is not conserved:

$$\begin{array}{rcl} uds & & uud \quad \bar{u}d \\ \Lambda^0 & = & p + \pi^- , \\ S = -1 & \neq & 0 + 0 \end{array} \quad (6.14)$$

The quark transformations necessary to accomplish these decay processes can be visualized with the help of Feynmann diagrams.

The omega-minus,  $\Omega^-$ , a baryon composed of three strange quarks, is a classic example of the need for the property called "color" in describing particles. Since quarks are fermions with spin 1/2, they must obey the Pauli Exclusion Principle and cannot exist in identical states. So with three strange quarks, the property which distinguishes them must be capable of at least three distinct values.

Conservation of strangeness is not in fact an independent conservation law, but can be viewed as a combination of the conservation of charge, isospin, and baryon number. It is often expressed in terms of hypercharge  $Y$ :

$$Y = S + B = 2(Q - I) , \quad (6.15)$$

where  $Q$ ,  $I$ ,  $S$  and  $B$  are the electric charge, isotopic spin, strangeness and baryon number, respectively.

Isospin and either hypercharge or strangeness are the quantum numbers often used to draw particle diagrams for the hadrons.

### 6.5.2 Charm

In 1974 a meson called the  $J/\Psi$  particle was discovered. With a mass of 3100 MeV, over three times that of the proton, this particle was the first example of another quark, called the charm quark. The  $J/\Psi$  is made up of a charm-anticharm quark pair.

The lightest meson which contains a charm quark is the  $D$  meson. It provides interesting examples of decay since the charm quark must be transformed into a strange quark by the weak interaction in order for it to decay.

One baryon with a charm quark, or having charm quantum number  $C = 1$ , is called a lambda with symbol  $\Lambda_c^+$ . It has a composition  $udc$  and a mass of 2281 MeV.

### 6.5.3 Beauty

In 1977, an experimental group at Fermilab led by Leon Lederman discovered a new resonance at 9.4 GeV which was interpreted as a bottom-antibottom quark pair and called the Upsilon meson,  $\Upsilon$ . From this experiment, the mass of the bottom quark is implied to be about 5 GeV. The reaction being studied was

$$p + N \rightarrow \mu^+ + \mu^- + X , \quad (6.16)$$

where  $N$  was a copper or platinum nucleus. The spectrometer had a muon-pair mass resolution of about 2%, which allowed them to measure an excess of events at 9.4 GeV. This resonance has been subsequently studied at other accelerators with a detailed investigation of the bound states of the bottom-antibottom meson. Studies of the  $B$ -meson have also been productive. The particles containing the bottom quark are assigned with the quantum number beauty,  $\tilde{B}$ .

### 6.5.4 Truth

Convincing evidence for the observation of the top quark was reported by Fermilab's Tevatron facility in April 1995. The evidence was found in the collision products of 0.9 TeV protons with equally energetic antiprotons in the proton-antiproton collider. The evidence involved analysis of trillions of

1.8 TeV proton-antiproton collisions. The Collider Detector Facility group had found 56 top candidates over a predicted background of 23 and the D0 group found 17 events over a predicted background of 3.8. The value for the top quark mass from the combined data of the two groups after the completion of the run was  $174.3 \pm 5.1$  GeV. This is over 180 times the mass of a proton and about twice the mass of the next heaviest fundamental particle, the  $Z^0$  vector boson at about 93 GeV.

By introduction of the quantum number truth,  $T$ , for the particles containing  $t$ -quark, the expression for the hypercharge in the Gell Mann-Nishijima formula obtains the form:

$$Y = B + S + C + \tilde{B} + T . \quad (6.17)$$

## 6.6 $C$ , $P$ and $T$ Transformations

Many of the profound ideas in nature manifest themselves as symmetries. A symmetry in a physical experiment suggests that something is conserved, or remains constant, during the experiment. So conservation laws and symmetries are strongly linked.

Three of the symmetries which usually, but not always, hold are those of charge conjugation ( $C$ ), parity ( $P$ ), and time reversal ( $T$ ):

- Charge conjugation ( $C$ ): reversing the electric charge and all the internal quantum numbers.
- Parity ( $P$ ): space inversion; reversal of the space coordinates, but not the time.
- Time reversal ( $T$ ): replacing  $t$  by  $-t$ . This reverses time derivatives like momentum and angular momentum.

### 6.6.1 Charge Conjugation

Classically, charge conjugation may seem like a simple idea: just replace positive charges by negative charges and vice versa. Since electric and magnetic fields have their origins in charges, you also must reverse these fields.

In quantum mechanical systems, charge conjugation has some further implications. It also involves reversing all the internal quantum numbers like those for lepton number, baryon number and strangeness. It does not affect mass, energy, momentum or spin.

Thinking of charge conjugation as an operator,  $C$ , then electromagnetic processes are invariant under the  $C$  operation since Maxwell's equations are invariant under  $C$ . This restricts some kinds of particle processes. Das and Ferbel proceed by defining a charge parity of  $\eta_C(\gamma) = -1$  for a photon since the  $C$  operation reverses the electric field. This constrains the electromagnetic decay of a neutral particle like the  $\pi^0$ . The decay of the  $\pi^0$  is:  $\pi^0 \rightarrow \gamma + \gamma$ . This implies that the charge parity or behavior under charge conjugation for a  $\pi^0$  is:

$$\eta_C(\pi^0) = \eta_C(\gamma)\eta_C(\gamma) = (-1)^2 = +1 . \quad (6.18)$$

Charge conjugation symmetry would imply that the  $\pi^0$  will not decay by  $\pi^0 \rightarrow \gamma$ , which we already know because it can't conserve momentum, but the decay  $\pi^0 \rightarrow 3\gamma$  can conserve momentum. This decay cannot happen because it would violate charge conjugation symmetry.

While the strong and electromagnetic interactions obey charge conjugation symmetry, the weak interaction does not. As an example, neutrinos are found to have intrinsic parities: neutrinos have left-handed parity and antineutrinos right-handed. Since charge conjugation would leave the spatial coordinates untouched, then if you operated on a neutrino with the  $C$  operator, you would produce a left-handed antineutrino. But there is no experimental evidence for such a particle; all antineutrinos appear to be right-handed. The combination of the parity operation  $P$  and the charge conjugation operation  $C$  on a neutrino do produce a right-handed antineutrino, in accordance with observation. So it appears that while beta decay does not obey parity or charge conjugation symmetry separately, it is invariant under the combination  $CP$ .

### 6.6.2 Parity

One of the conservation laws which applies to particle interactions is associated with parity. Parity involves a transformation that changes the algebraic sign of the coordinate system. Parity is an important idea in quantum mechanics because the wavefunctions which represent particles can behave in

different ways upon transformation of the coordinate system which describes them. The parity transformation changes a right-handed coordinate system into a left-handed one or vice versa. Two applications of the parity transformation restores the coordinate system to its original state.

Quarks have an intrinsic parity which is defined to be +1 and for an antiquark parity = -1. Nucleons are defined to have intrinsic parity +1. For a meson with quark and antiquark with antiparallel spins ( $s = 0$ ), then the parity is given by

$$P = P_q P_{\bar{q}} (-1)^l, \quad (6.19)$$

where  $l$  is orbital angular momentum.

The meson parity is given by

$$P = -(-1)^l = (-1)^{l+1}. \quad (6.20)$$

The lowest energy states for quark-antiquark pairs (mesons) will have zero spin and negative parity and are called pseudoscalar mesons. The nine pseudoscalar mesons can be shown on a meson diagram. One kind of notation for these states indicates their angular momentum and parity

$$j^P = 0^-. \quad (6.21)$$

Excited states of the mesons occur in which the quark spins are aligned, which with zero orbital angular momentum gives  $j = 1$ . Such states are called vector mesons,

$$j^P = 1^-. \quad (6.22)$$

The vector mesons have the same spin and parity as photons.

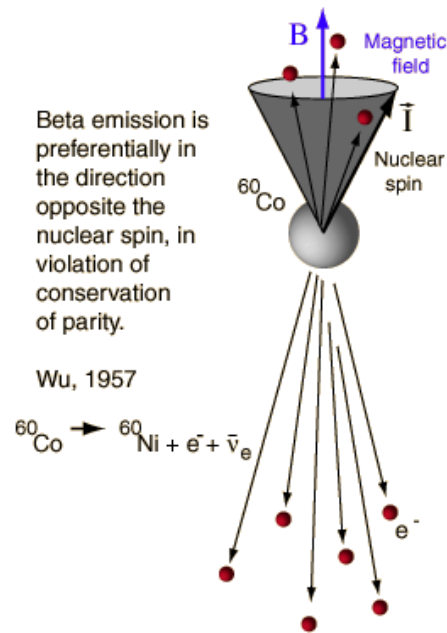
All neutrinos are found to be "left-handed", with an intrinsic parity of -1 while antineutrinos are right-handed, parity = +1.

### Non-conservation of Parity

The electromagnetic and strong interactions are invariant under the parity transformation. It was a reasonable assumption that this was just the way nature behaved, oblivious to whether the coordinate system was right-handed or left-handed. But for several years physicists had puzzled over the

decay of the neutral kaons, which had equal mass but decayed to products of opposite parity. In 1956, Lee and Yang predicted the non-conservation of parity in the weak interaction. Their prediction was quickly tested when Wu and collaborators studied the beta decay of Cobalt-60 in 1957.

By lowering the temperature of cobalt atoms to about 0.01K, Wu was able to "polarize" the nuclear spins along the direction of an applied magnetic field. The directions of the emitted electrons were then measured. Equal numbers of electrons should be emitted parallel and antiparallel to the magnetic field if parity is conserved, but they found that more electrons were emitted in the direction opposite to the magnetic field and therefore opposite to the nuclear spin.



This and subsequent experiments have consistently shown that a neutrino always has its intrinsic angular momentum (spin) pointed in the direction opposite its velocity. It is called a left-handed particle as a result. Anti-neutrinos have their spins parallel to their velocity and are therefore right-handed particles. Therefore we say that the neutrino has an intrinsic chirality.

The idea that nature at a very fundamental level can tell the difference

between "left-handed" and "right-handed" systems is a radical one. It was thought for a time that the combination of the parity operation ( $P$ ) and "charge conjugation" ( $C$ ) was an inviolate conservation law ( $CP$  invariance). But the study of the Kaon decay in 1964 showed a violation of  $CP$ .

### 6.6.3 Time Reversal

In simple classical terms, time reversal just means replacing  $t$  by  $-t$ , inverting the direction of the flow of time. Reversing time also reverses the time derivatives of spatial quantities, so it reverses momentum and angular momentum. Newton's second law is quadratic in time and is invariant under time reversal. Its invariance under time reversal holds for either gravitational or electromagnetic forces.

Very sensitive experimental tests have been done to put upper bounds on any violation of time-reversal symmetry. One experiment described by Das and Ferbel is the search for a dipole moment for the neutron. Even though the neutron is neutral, it is viewed as made up of charged quarks and therefore could conceivably have a dipole moment. Experimental evidence is consistent with zero dipole moment, so time reversal symmetry seems to hold in this case.

The small violation of  $CP$  symmetry suggests some departure from  $T$  symmetry in some weak interaction process since  $CPT$  invariance seems to be on very firm ground.

### 6.6.4 CP Invariance

The strong and electromagnetic interactions leave systems invariant under any of the three operations applied alone, but the weak interaction does not. The beta decay of cobalt-60 established the violation of parity in 1957, and led to our understanding that the weak interaction violates both charge conjugation and parity invariance. However, the weak interaction does appear to leave systems invariant under the combination  $CP$ . Examination of the case of the neutrino is instructive at this point. The parity operation on a neutrino would leave its spin in the same direction while reversing space coordinates.

$$P|\nu_L\rangle \rightarrow |\nu_L\rangle, \quad P|\nu_R\rangle \rightarrow |\nu_R\rangle. \quad (6.23)$$



Neither of these things is observed to happen in nature; neutrinos are always left-handed, anti-neutrinos always right-handed. But if you add the charge conjugation operation, the result of the combined operation gives you back the original particle.

$$CP|\nu_L\rangle \rightarrow C|\nu_L\rangle = |\nu_L\rangle, \quad CP|\nu_R\rangle \rightarrow C|\nu_R\rangle = |\nu_R\rangle. \quad (6.24)$$

*CP* invariance was thought to be a general conservation principle until the details of the neutral kaon decay process were examined by Cronin and Fitch. After intense study over many years, the consensus is that *CP* is violated by a small amount. In 2001 *CP* violation was confirmed in *B*-meson decay. It is thought possible by some investigators that in *CP* violation is to be found the reason for the vast excess of matter over antimatter in the universe.

### 6.6.5 *CPT* Theorem

Conservation laws for parity, isospin, and strangeness have been developed by detailed observation of particle interactions. The combination of charge conjugation (*C*), parity (*P*) and time reversal (*T*) is considered to be a fundamental symmetry operation - all physical particles and interactions appear to be invariant under this combination. Called *CPT* invariance, this symmetry plumbs the depths of our understanding of nature.

On the theoretical side, *CPT* invariance has received a great deal of attention. Georg Ludens, Wolfgang Pauli and Julian Schwinger independently showed that invariance under Lorentz transformations implies *CPT* invariance. *CPT* invariance itself has implications which are at the heart of our understanding of nature and which do not easily arise from other types of considerations.

- Integer spin particles obey Bose-Einstein statistics and half-integer spin particles obey Fermi-Dirac statistics. Operators with integer spins must be quantized using commutation relations, while anticommutation relations must be used for operators with half integer spin;
- Particles and antiparticles have identical masses and lifetimes. This arises from *CPT* invariance of physical theories;
- All the internal quantum numbers of antiparticles are opposite to those of the particles.

**Exercise 6.1:** Why does the proton have a parity while the muon does not?

**Exercise 6.2:** A state containing only one strange particle: (a) can decay into a state of zero strangeness; (b) can be created strongly from a state of zero strangeness; (c) cannot exist.

**Exercise 6.3:** For each of the following decays state a conservation law that forbids it:  $n \rightarrow p + e^-$ ;  $n \rightarrow \pi^+ + e^-$ ;  $n \rightarrow p + \pi^-$  and  $n \rightarrow p + \gamma$ .

**Exercise 6.4:** Check that the Gell-Mann-Nishijima formula works for the quarks  $u$ ,  $d$  and  $s$ . Using the table of quark properties, and the quark isospin assignments, deduce the generalized Gell-Mann-Nishijima formula expressing  $Q$  in terms of  $B$ ,  $I_3$ ,  $S$ ,  $C$ ,  $B$  and  $T$ .

**Exercise 6.5:** Is the neutrino an eigenstate of  $P$ ? If so, what is its intrinsic parity?

**Exercise 6.6:** Give a non-trivial (rate greater than 5%) decay mode for each particle in the following list:  $n$ ,  $\pi^+$ ,  $\rho^0$ ,  $K^0$ ,  $\Lambda^0$ ,  $\Delta^{++}$ ,  $\mu^-$ ,  $\Omega^-$ ,  $J/\Psi$ .

**Exercise 6.7:** List all of the known leptons. How does  $\mu^+$  decay? Considering this decay and the fact that  $\nu_\mu + n \rightarrow e^- + p$  is found to be forbidden, discuss possible lepton quantum number assignments that satisfy additive quantum number conservation laws. How could  $\nu_\mu$  produce a new charged "heavy lepton"?

**Exercise 6.8:** Having 4.5 GeV free energy, what is the most massive isotope one could theoretically produce from nothing?

**Exercise 6.9:** Which conservation law forbids the decay  $K \rightarrow \pi\gamma$ ?

**Exercise 6.10:** There are no known mesons of electric charge two. Can you give a simple explanation of this?

**Exercise 6.11:** Explain how the parity of pion can be measured by observation of the polarizations of the photons in  $\pi^0 \rightarrow \gamma\gamma$ .

**Exercise 6.12:** To a very high accuracy, the cross section for  $e^-p$  scattering equals the cross section for  $e^+p$  scattering. Is this equality a consequence of a conservation law? If so, which one? If not, explain the observed equality.

## Part III

# Lecture – Quantum Field Theory



# Chapter 7

## Quantum Fields

### 7.1 Primary and Secondary Quantizations

In this lecture our main goal is to formulate a relativistic quantum mechanical theory of interactions. We start with the fundamental equation of quantum mechanics, Schroedinger's equation,

$$H\Psi = i\hbar\frac{\partial\Psi}{\partial t} . \tag{7.1}$$

We know that for a non-interacting, non-relativistic particle,

$$H = \frac{\vec{p}^2}{2m} = -\frac{\hbar}{2m}\vec{\nabla}^2 , \tag{7.2}$$

so

$$-\frac{\hbar}{2m}\vec{\nabla}^2\Psi = i\hbar\frac{\partial\Psi}{\partial t} . \tag{7.3}$$

Of course, in this case  $\Psi$  is a *scalar field*, and therefore only has one state. So, it describes a spin-0 particle. Since  $\Psi$  does not have any spacetime indices, it also transforms trivially under the Lorentz group.

Notice, however, that we have a fundamental barrier in making a relativistic theory – the spatial derivative in (7.3) acts quadratically ( $\vec{\nabla}^2$ ), whereas the time derivative is linear. Clearly, treating space and time differently in this way is unacceptable for a relativistic theory. That is a hint of a much more

fundamental problem with quantum mechanics; space is always treated as an operator, but time is always treated as a parameter. This fundamental asymmetry is what ultimately prevents a straightforward generalization to relativistic quantum theory.

To fix this problem, we have two choices: either promote time to an operator along with space, or demote space back to a parameter and quantize in a new way.

The first option would result in the Hermitian operators  $\hat{X}$ ,  $\hat{Y}$ ,  $\hat{Z}$ , and  $\hat{T}$ . It turns out that this approach is very difficult and less useful as far as building a relativistic quantum theory. So, we will take the second option.

In demoting position to a parameter along with time, we obviously have sacrificed the operators which we imposed commutation relations on to get a 'quantum' theory in the first place. Because we obviously can't impose commutation relations on parameters (because they are scalars), quantization appears impossible. So, we are going to have to make a fairly radical reinterpretation.

Rather than letting the coordinates be Hermitian operators that act on the state in the Hilbert space representing a particle, we now interpret the particle as the Hermitian operator, and this operator (or particle) will be parameterized by the spacetime coordinates. The physical state that the particle operators act on is then the vacuum itself,  $|0\rangle$ . So, whereas before you acted on the 'electron'  $|\Psi\rangle$  with the operator  $\hat{x}$ , now the 'electron' (parameterized by  $x$ )  $\Psi(x^\mu)$  acts on the vacuum  $|0\rangle$ , creating the state  $\Psi(x^\mu)|0\rangle$ . In other words, the operator representing an electron excites the vacuum (empty space) resulting in an electron. We will see that all quantum fields contain appropriate raising and lowering operators to do just this.

This approach, where the quantum mechanical entities are no longer the coordinates acting on the fields, but the fields themselves, is called *Quantum Field Theory* (QFT).

So, whereas before, quantization was defined by imposing commutation relations on the coordinate operators  $[x, p] \neq 0$  (*primary quantization*), we now quantize by imposing commutation relations on the field operators,  $[\Psi_1, \Psi_2] \neq 0$ , what sometimes is called as *secondary quantization*.

Because we must still write down the equations of motion which govern the

dynamics of the fields, we want to remind some classical equations governing the fields we want to work with (see Chapt. 2).

## 7.2 Quantization of Spin-0 Fields

As we said above, Schrodinger's equation (7.3) describes the time evolution of a spin-0 field, or a scalar field. Generalizing to higher spins will come later. Now, we see how to make this description relativistic.

The most obvious guess for a relativistic form is to simply plug in the standard relativistic Hamiltonian

$$H = \sqrt{\vec{p}^2 c^2 + m^2 c^4} . \quad (7.4)$$

Note that the Taylor expansion,

$$\sqrt{1 + x^2} \approx 1 + \frac{x^2}{2} , \quad (x \ll 1) \quad (7.5)$$

gives the Hamiltonian for  $\vec{p}^2 \ll c^2$ ,

$$H \approx mc^2 + \frac{\vec{p}^2}{2m} , \quad (7.6)$$

which is the standard non-relativistic form (plus a constant) we'd expect from a low speed limit.

Plugging (7.4) into (7.3), we have

$$i\hbar \frac{\partial \phi}{\partial t} = \sqrt{-\hbar^2 c^2 \vec{\nabla}^2 + m^2 c^4} \phi . \quad (7.7)$$

But there are two problems with this:

1. The space and time derivatives are still treated differently, so this is inadequate as a relativistic equation;
2. Taylor expanding the square root will give an infinite number of derivatives acting on  $\phi$ , making this theory non-local.

One solution is to square the operator on both sides, giving the *Klein Gordon* equation:

$$-\hbar^2 \frac{\partial^2 \phi}{\partial t^2} = \left( -\hbar^2 c^2 \vec{\nabla}^2 + m^2 c^4 \right) \phi \Rightarrow \left( -\partial^0 \partial_0 + \vec{\nabla}^2 - \frac{m^2 c^2}{\hbar^2} \right) \phi = 0 . \quad (7.8)$$

If we choose the so called 'natural units' or 'God units', where  $c = \hbar = 1$ , we have

$$(\partial^2 - m^2) \phi = 0 . \quad (7.9)$$

The equation is nothing more than an operator version of the standard relativistic relation

$$E^2 = m^2 c^4 + \vec{p}^2 c^2 . \quad (7.10)$$

Note that because we will be quantizing fields and not coordinates, there is absolutely nothing 'quantum' about the Klein-Gordon equation. It is, at this point, merely a relativistic wave equation for a classical, spinless, non-interacting field (see Sec. 2.2).

Finally, we note one major problem with the Klein-Gordon equation. When we squared the Hamiltonian (7.4) to get

$$H^2 = m^2 c^4 + \vec{p}^2 c^2 , \quad (7.11)$$

the energy eigenvalues became

$$E = \pm \sqrt{m^2 c^4 + \vec{p}^2 c^2} . \quad (7.12)$$

It appears that we have a negative energy eigenvalue! Obviously this is unacceptable in a physically meaningful theory, because negative energy means that we don't have a true vacuum, and therefore a particle can cascade down forever, giving off an infinite amount of radiation.

We will see that this problem plagues the spin-1/2 particles as well, so we wait to talk about the solution until then.

### 7.3 Quantization of Spin 1/2 Fields

Finding equation (7.9) was easy because scalar fields have no spacetime indices and no spinor indices. A particle of spin 1/2 however, will have two



complex components, one for spin  $+1/2$ , and the other for spin  $-1/2$  (see Sec. 2.5). So, we describe such a particle as the two-component Spinor,

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \quad (7.13)$$

where  $\psi_1$  and  $\psi_2$  are both  $\in \mathbb{C}$ . So, we want some differential operator in the form of  $2 \times 2$  matrices to act on such a field to form the equation of motion.

Following Dirac's approach, he reasoned that given such a  $2 \times 2$  operator, the equation of motion should somehow 'imply' the Klein-Gordon equation (which merely makes the theory relativistic). So his goal (and our goal) is to find an equation with a  $2 \times 2$  matrix differential operator acting on  $\psi$  that results in (7.9).

Dirac's approach was to find an operator of the form

$$\mathcal{D} = \gamma^\mu \partial_\mu = \gamma^0 \partial_0 + \gamma^1 \partial_1 + \gamma^2 \partial_2 + \gamma^3 \partial_3, \quad (7.14)$$

where the  $\gamma$ 's are  $2 \times 2$  matrices, and the equation of motion is then

$$\mathcal{D}\psi = -im\psi. \quad (7.15)$$

The challenge is in finding the appropriate  $2 \times 2$   $\gamma$ -matrices. Dirac reasoned that, in order to be properly relativistic, operating twice with  $\mathcal{D}$  should give the Klein-Gordon equation. In other words,

$$\begin{aligned} \mathcal{D} = -im &\Rightarrow \mathcal{D}\mathcal{D}\psi = -im\mathcal{D}\psi, \\ &\Rightarrow \gamma^\mu \partial_\mu \gamma^\nu \partial_\nu \psi = -im(-im\psi), \\ &\Rightarrow \gamma^\mu \gamma^\nu \partial_\mu \partial_\nu \psi = -m^2 \psi, \\ &\Rightarrow (\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu + m^2)\psi = 0. \end{aligned} \quad (7.16)$$

This will yield the Klein-Gordon equation if

$$\gamma^\mu \gamma^\nu = -\eta^{\mu\nu} \mathbb{I}. \quad (7.17)$$

Or, using the symmetry of the sum in (7.16), it will yield the Klein-Gordon equation if we demand

$$\frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) = -\eta^{\mu\nu} \mathbb{I}, \quad (7.18)$$

Consider

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = -2\eta^{\mu\nu} \mathbb{I} \quad (7.19)$$

If the  $\gamma$ -matrices satisfy (7.19), then (7.16) gives

$$(\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu + m^2)\psi = 0 \quad \Rightarrow \quad (-\eta^{\mu\nu} \partial_\mu \partial_\nu + m^2)\psi = 0, \quad (7.20)$$

which is exactly the Klein-Gordon equation (7.9).

So, we have the Dirac equation

$$(\mathcal{D} + im)\psi = 0, \quad (7.21)$$

but we still have a problem. Namely, there does not exist a set of  $2 \times 2$  matrices that solve (7.19). Nor does there exist a set of  $3 \times 3$  matrices. The smallest possible size where this is possible is  $4 \times 4$ . Obviously, if we want to describe a spin-1/2 particle with exactly 2 spin states, using 4 spin components does not seem right. But, we will accept the necessity of  $4 \times 4$  Dirac matrices and move on.

Instead of using  $(\psi_1, \psi_2)$ , we will define the two 2-dimensional spinors

$$\psi_L \equiv \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \quad \text{and} \quad \psi_R \equiv \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} \quad (7.22)$$

and the 4-component spinor

$$\psi \equiv \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \quad (7.23)$$

Now it is possible to solve (7.19). Such a problem is actually very familiar to algebraists, and we will not delve into the details of how this is done. Instead, we merely state one solution (there are many, up to a similarity transformation). We define the  $4 \times 4$  matrices,

$$\gamma^i = \begin{pmatrix} 0 & -\sigma^i \\ \sigma^i & 0 \end{pmatrix} \quad \text{and} \quad \gamma^0 = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix}, \quad (7.24)$$

where  $\sigma^0$  is the  $2 \times 2$  identity matrix, and  $\sigma^i$  are the Pauli spin matrices. It should be no surprise that they show up in attempting to describe spin-1/2 particles. What is interesting is that we did not assume them – we derived them using (7.19).

Now that we have an explicit form of the Dirac gamma matrices, we can write out (7.21) explicitly:

$$\begin{pmatrix} 0 & 0 & \partial_0 - \partial_3 & -\partial_1 + i\partial_2 \\ 0 & 0 & -\partial_1 - i\partial_2 & \partial_0 + \partial_3 \\ \partial_0 + \partial_3 & \partial_1 - i\partial_2 & 0 & 0 \\ \partial_1 + i\partial_2 & \partial_0 - \partial_3 & 0 & 0 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = -im \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \quad (7.25)$$

Or, in terms of  $\psi_L$  and  $\psi_R$ ,

$$\begin{aligned} i\bar{\sigma}^\mu \partial_\mu \psi_R &= +m\psi_L, \\ i\sigma^\mu \partial_\mu \psi_L &= +m\psi_R, \end{aligned} \tag{7.26}$$

where we have defined the 4-vectors

$$\sigma^\mu = (\sigma^0, \sigma^1, \sigma^2, \sigma^3) \quad \text{and} \quad \bar{\sigma}^\mu = (\sigma^0, -\sigma^1, -\sigma^2, -\sigma^3). \tag{7.27}$$

### 7.3.1 The Dirac Sea Interpretation of Antiparticles

Initially, it may seem that the impossibility of finding a  $2 \times 2$  matrix solution to (7.19) means that we can't have fields with 2 spinor states. However, we aren't limited to scalars and spacetime 4-component spinors. We can also have two fields,  $\psi_L$  and  $\psi_R$ , which can be paired together to form two spin-1/2 fields in a 4-component spinor  $(\psi_L, \psi_R)$ . So, Dirac was faced with the challenge of both interpreting this, while at the same time dealing with the negative energy states mentioned in section 7.2.

Dirac's solution, though today abandoned, was brilliant enough to mention. He suggested that because spin-1/2 particles obey the *Pauli Exclusion Principle*, there could be an infinite number of particles already in the negative energy levels, and so they are already occupied, preventing any more particles from falling down and giving off infinite energy. Thus, the negative energy problem was solved.

Furthermore, he said that it is possible for one of the particles in this infinite negative sea to be excited and jump up into a positive energy state, leaving behind a hole. This would appear to us, experimentally, as a particle with the same mass, but the opposite charge. He called such particles *Antiparticles*. For example, the antiparticle of the electron is the antielectron, or the positron (same mass, opposite charge). The positron is not a particle in the same sense as the electron, but rather is a hole in an infinite sea of electrons. And where this negative charge is missing, all that is left is a hole which appears as a positively charged particle.

So,  $\psi_L$  describes a particle, and due to the infinite sea of negative particles, there can always be a hole, which will be described by  $\psi_R$ . Everything about this worked out mathematically, and when antiparticles were detected about 5 years after Dirac's prediction of them, it appeared that Dirac's suggestion was correct.

However, there were two major problems with Dirac's idea, and they ultimately proved fatal to the 'Dirac Sea' interpretation:

1. This theory, which was supposed to be a theory of single particles, now requires an infinite number of them;
2. Particles like photons, pions, mesons, or Klein-Gordon scalars don't obey the Pauli Exclusion Principle, but still have negative energy states, and therefore Dirac's argument doesn't work.

However, his labeling them 'antiparticles' has stuck, and we therefore still refer to the right-handed part of the spin-1/2 field,  $\psi_R$ , as the antiparticle, whereas the left-handed part,  $\psi_L$ , is still the particle.

For these reasons, we must have some other way of understanding the existence of the antiparticles.

## 7.4 The QFT Interpretation of Antiparticles

In presenting the problem of negative energy states, we have been somewhat intentionally sloppy. To take stock, we have two equations of motion: the Klein-Gordon (7.9) for scalar/spin-0 fields, and the Dirac equation (7.21) for spin-1/2 particles.

In our discussion of negative energy states, we were 'pretending' that the  $\psi$ 's and  $\phi$ 's have 'states' with negative energy. But, as we said in section 7.1, QFT offers a different interpretation of the fields. Namely, the fields are not states – they are operators. And consequently they can't have energy. A state is made by acting on the vacuum with either of the operators  $\phi$  or  $\psi$ , and then the state  $\phi|0\rangle$  or  $\psi|0\rangle$  has some positive energy.

So, QFT allows us to see the antiparticle as a real, actual particle, rather than the absence of a particle. And, we do not need the conceptually difficult idea of an infinite sea of negative energy particles. The vacuum  $|0\rangle$ , with no particles in it, is now our state with the lowest possible energy level.

How exactly  $|0\rangle$  works will become clearer when we quantize. The point to be understood for now is that QFT solves the problem of negative energy by reinterpreting what is a state and what is an operator. The fields  $\phi$  and  $\psi$

are operators, not states, and therefore they do not have energy associated with them (any more than the operator  $\hat{x}$  or  $\hat{p}_x$  did in non-relativistic quantum mechanics). So, without any problems of negative energy, we merely accept that nature, due to relativity, demands that particles come in particle/antiparticle pairs, and we move on.

## 7.5 Dirac Equation with Electromagnetic Field

In Sec. 2.1 we introduced the Lagrangian for an electromagnetic field. Our goal now is to find a Lagrangian that describes the electromagnetic field and a spin-1/2 particle that couples to the electromagnetic field, and additionally the interaction between them. We start by writing down a Lagrangian without any interaction, which is simply the sum of the two terms,

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{EM} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu , \quad (7.28)$$

where  $J^\mu = \bar{\psi}\gamma^\mu\psi$  denotes the global  $U(1)$  current.

Because the Dirac part has no terms in common with the electromagnetic part, the equations of motion and the conserved quantities for both  $\psi$  and  $A^\mu$  will be exactly the same, as if the other weren't present at all. In other words, both fields go about their way as if the other weren't there – there is no interaction in this theory. Because this makes for a boring universe (and horrible phenomenology), we need to find some way of coupling the two fields together to produce some sort of interaction.

Interaction is introduced in a physical theory by adding the *Interaction Term* to the Lagrangian. So, the final Lagrangian will have the form

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{EM} + \mathcal{L}_{int} . \quad (7.29)$$

We do this by coupling the electromagnetic field  $A^\mu$  to the current resulting from the local  $U(1)$  symmetry in  $\mathcal{L}_D$ . In other words, our interaction term will be proportional to  $A^\mu j_\mu$ .

So, adding a constant of proportionality  $q$  (which we will see has the physical interpretation of a coupling constant, weighting the probability of an interaction to take place, or equivalently the physical interpretation of electric

charge), our Lagrangian is now

$$\begin{aligned}\mathcal{L} &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu - qj^\mu A_\mu = \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - (J^\mu + q\bar{\psi}\gamma^\mu\psi)A_\mu .\end{aligned}\quad (7.30)$$

Notice that  $\mathcal{L}$  is still invariant under the global  $U(1)$ , and that the Lagrangians (7.28) and (7.30) are the same except for a shift of the current,

$$J^\mu \rightarrow J^\mu + qj^\mu .\quad (7.31)$$

The fact that  $J^\mu$  has shifted in (7.30) means that the spin-1/2 particle in this theory contributes to the field, which is what we would expect it to do.

If we set  $q = e$ , the electric charge, this Lagrangian becomes upon quantization the Lagrangian of Quantum Electrodynamics (*QED*), which to date makes the most accurate experimental predictions ever. Below we shall write this lagrangian in more fundamental form.

**Exercise 7.1:** Derive the equation for a massive spin-1 field, find corresponding propagator and the conditions on the polarization vector for its plane-wave solution.

**Exercise 7.2:** Construct a Lagrangian for scalar QED, with a complex scalar fields and a photon field. Identify the interaction terms and derive the Feynman rules.

**Exercise 7.3:** The Pauli-Lubanski vector is defined by

$$W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}M^{\nu\rho}p^\sigma ,$$

where  $p^\sigma = -i\partial^\sigma$  is the linear momentum and

$$M^{\nu\rho} = \frac{1}{2}\sigma^{\nu\rho} + i(x^\nu\partial^\rho - x^\rho\partial^\nu)$$

is the angular momentum. For spinor fields the spin momentum here is defined as:

$$\sigma^{\nu\rho} = \frac{i}{2}[\gamma^\nu, \gamma^\rho] .$$

Find the expression for  $W^2\psi(x)$ , where  $\psi(x)$  is a solution of the Dirac equation.

**Exercise 7.4:** Find the Lagrangian density of a spinless Schrödinger field.

## Chapter 8

# Canonical Quantization

In quantum mechanics (not QFT), quantization is done by taking certain dynamical quantities and making use of the *Heisenberg Uncertainty Principle*. Normally we take position  $\vec{x}$  and momentum  $\vec{p}$  and, according to Heisenberg, the measurement of the particle's position will affect its momentum and vice-versa (*primary quantization*).

To make this more precise, we promote  $x$  and  $p$  from merely being variables to being Hermitian operators  $\hat{x}$  and  $\hat{p}$  (which can be represented by matrices) acting on some vector space. Calling a vector in this space  $|\psi\rangle$ , physically measurable quantities (like position or momentum) become the eigenvalues of the operators  $\hat{x}$  and  $\hat{p}$ ,

$$\hat{x}|\psi\rangle = x|\psi\rangle, \quad \hat{p}|\psi\rangle = p|\psi\rangle. \quad (8.1)$$

Heisenberg Uncertainty says that measuring  $x$  will affect the value of  $p$ , and vice-versa. It is the act of measuring which enacts this effect. It is not an engineering problem in the sense that there is no better measurement technique which would undo this. It is a fundamental fact of quantum mechanics (and therefore the universe) that measurement of one variable affects another.

So, if we measure  $x$  (using  $\hat{x}$ ) and then  $p$  (using  $\hat{p}$ ), we will in general get different values for both than if we measured  $p$  and then  $x$ . More mathematically,  $\hat{x}\hat{p} \neq \hat{p}\hat{x}$ . Put another way,

$$[\hat{x}, \hat{p}] \equiv \hat{x}\hat{p} - \hat{p}\hat{x} \neq 0. \quad (8.2)$$

For reasons learned in a quantum course, the actual relation is

$$[\hat{x}, \hat{p}] = i\hbar, \quad (8.3)$$

where  $\hbar$  is Planck's Constant. We call (8.3) the *Canonical Commutation Relation*, and it is this structure which allows us to determine the physical structure of the theory.

More generally, we choose some set of operators that all commute with each other, and then label a physical state by its eigenvectors. For example  $\hat{x}$ ,  $\hat{y}$  and  $\hat{z}$  all commute with each other, so we may label a physical state by its eigenvectors  $|\psi_r\rangle = |x, y, z\rangle$ . Or, because  $\hat{p}_x$ ,  $\hat{p}_y$ , and  $\hat{p}_z$  all commute, we may call the state  $|\psi_p\rangle = |p_x, p_y, p_z\rangle$ . We may also include some other values like spin and angular momentum, to have (for example)  $|\psi\rangle = |x, y, z, s_z, L_z, \dots\rangle$ .

As discussed in section 7.1, when we make the jump to QFT, the fields are no longer the states but the operators. We are therefore going to impose commutation relations on the fields, not on the coordinates.

Furthermore, whereas before the states were eigenvectors of the coordinate operators, we now will expand the fields in terms of the eigenvectors of the Hamiltonian.

## 8.1 Canonical Quantization of Scalar Fields

We begin with the Klein-Gordon Lagrangian, which is modified by an arbitrary constant  $\Omega$ :

$$\mathcal{L}_{KG} = -\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2 + \Omega. \quad (8.4)$$

Note that  $\Omega$  has absolutely no affect whatsoever on the physics.

Quantization then comes about by defining the field momentum and Hamiltonian,

$$\begin{aligned} \Pi &= \frac{\partial\mathcal{L}}{\partial\dot{\phi}(x)} = \dot{\phi}, \\ \mathcal{H} &= \Pi\dot{\phi} - \mathcal{L} = \frac{1}{2}\Pi^2 + \frac{1}{2}(\vec{\nabla}\phi)^2 + \frac{1}{2}m^2\phi^2 - \Omega. \end{aligned} \quad (8.5)$$



Now, using the canonical commutation relations (8.3) as guides, we impose

$$\begin{aligned} [\phi(t, \vec{x}), \phi(t', \vec{x}')] &= 0, \\ [\Pi(t, \vec{x}), \Pi(t', \vec{x}')] &= 0, \\ [\phi(t, \vec{x}), \Pi(t', \vec{x}')] &= i\delta(t - t')\delta(\vec{x} - \vec{x}') \end{aligned} \quad (8.6)$$

(where we have set  $\hbar = 1$ ).

We can see more clearly what this means if we expand the solutions of the Klein-Gordon equation. One solution is plane waves,  $e^{i\vec{k}\cdot\vec{x} \pm i\omega t}$ , where

$$\omega = +\sqrt{\vec{k}^2 + m^2}, \quad (8.7)$$

and  $\vec{k}$  is the standard wave vector.

So, we write the field  $\phi$  as

$$\phi(t, \vec{x}) = \int \frac{d^3\vec{k}}{f(\vec{k})} \left[ a(\vec{k})e^{i\vec{k}\cdot\vec{x} - i\omega t} + b(\vec{k})e^{i\vec{k}\cdot\vec{x} + i\omega t} \right], \quad (8.8)$$

where  $f(x)$  is a redundant function which we have included for later convenience. For now, both  $a(\vec{k})$  and  $b(\vec{k})$  are merely arbitrary coefficients (integration constants) used to expand  $\phi(t, \vec{x})$  in terms of individual solutions.

We demand that  $\phi(t, \vec{x})$  be Hermitian. This requires

$$\phi^\dagger = \phi \quad \Rightarrow \quad \phi^* = \phi \quad \Rightarrow \quad b^*(\vec{k}) = a(-\vec{k}). \quad (8.9)$$

Then, changing the sign of the integration variable  $\vec{k}$  on the second term in the integral allows us to use 4-vector notation, so

$$\phi(x) = \int \frac{d^3\vec{k}}{f(\vec{k})} \left[ a(\vec{k})e^{ik\cdot x} + a^*(\vec{k})e^{-ik\cdot x} \right], \quad (8.10)$$

where  $k \cdot x = k^\mu x_\mu$ .

Now notice that the integration measure,  $d^3\vec{k}$ , is not invariant under Lorentz transformations (because it integrates over the spatial part but not over the time part). We therefore choose  $f(\vec{k})$  to restore Lorentz invariance.

We know that the measure  $d^3k$  would be invariant, as would  $\delta$ -functions and  $\Theta$  (step) functions. So, consider the invariant combination

$$d^4k\delta(k^2 + m^2)\Theta(k^0). \quad (8.11)$$

The  $\delta$ -function merely requires that relativity hold,  $k^2 + m^2$  is simply the relativistic relation (7.4), and the  $\Theta$ -function preserves causality. So this is a physically acceptable Lorentz invariant integration measure.

Recall the general  $\delta$ -function identity,

$$\int_{-\infty}^{\infty} dx \delta[g(x)] = \sum_i \frac{1}{|dg(x)/dx|_{x=x_i}} , \quad (8.12)$$

where the  $x_i$ 's are the zeros of the function  $g(x)$ . We can do the  $k^0$  integral over measure (8.11), and using the fact that the zeros of

$$k^2 + m^2 = \vec{k}^2 - k^0 k_0 + m^2 \quad (8.13)$$

in terms of  $k^0$  are  $k^0 k_0 = \vec{k}^2 + m^2 = \omega^2$ , we get

$$\int d^3 \vec{k} dk^0 \delta(k^2 + m^2) \Theta(k^0) = \int \frac{d^3 \vec{k}}{2\omega} . \quad (8.14)$$

So, adding a factor of  $(2\pi)^3$  for later convenience, we take our invariant measure to be  $d^3 \vec{k}/(2\pi)^3 2\omega$ . So finally,

$$\phi(x) = \int \widetilde{d\vec{k}} \left[ a(\vec{k}) e^{ik \cdot x} + a^*(\vec{k}) e^{-ik \cdot x} \right] , \quad (8.15)$$

where we have defined

$$\widetilde{d\vec{k}} \equiv \frac{d^3 \vec{k}}{(2\pi)^3 2\omega} . \quad (8.16)$$

The commutation relations we defined in (8.6) will now hold provided we impose

$$\begin{aligned} [a(\vec{k}), a(\vec{k}')] &= 0 , \\ [a^\dagger(\vec{k}), a^\dagger(\vec{k}')] &= 0 , \\ [a(\vec{k}), a^\dagger(\vec{k}')] &= (2\pi)^3 2\omega \delta^3(\vec{k} - \vec{k}') . \end{aligned} \quad (8.17)$$

We are using  $\dagger$  instead of  $\star$  to emphasize that, in the quantum theory, we are talking about Hermitian operators. The operators  $a(\vec{k})$  and  $a^\dagger(\vec{k})$  are scalars, so in this case  $a^\star = a^\dagger$ .

Furthermore, we can write the Hamiltonian  $H$  in terms of (8.15):

$$\begin{aligned}
H &= \int d^3x \mathcal{H} = \int d^3x \left( \frac{1}{2} \Pi^2 + \frac{1}{2} (\vec{\nabla} \phi)^2 + \frac{1}{2} m^2 \phi^2 - \Omega \right) = \\
&= - \int d^3x \Omega + \frac{1}{2} \int \widetilde{dk} \widetilde{dk}' d^3x \left[ \left( -i\omega a(\vec{k}) e^{ik \cdot x} + i\omega a^*(\vec{k}) e^{-ik \cdot x} \right) \times \right. \\
&\quad \times \left( -i\omega' a(\vec{k}') e^{ik' \cdot x} + i\omega' a^*(\vec{k}') e^{-ik' \cdot x} \right) + \\
&\quad + \left( i\vec{k} a(\vec{k}) e^{ik \cdot x} - i\vec{k} a^*(\vec{k}) e^{-ik \cdot x} \right) \left( i\vec{k}' a(\vec{k}') e^{ik' \cdot x} - i\vec{k}' a^*(\vec{k}') e^{-ik' \cdot x} \right) + \\
&\quad \left. + m^2 \left( a(\vec{k}) e^{ik \cdot x} + a^*(\vec{k}) e^{-ik \cdot x} \right) \left( a(\vec{k}') e^{ik' \cdot x} + a^*(\vec{k}') e^{-ik' \cdot x} \right) \right] \\
&= \frac{1}{2} \int \widetilde{dk} \widetilde{dk}' d^3x \left[ \left( -\omega\omega' a(\vec{k}) a(\vec{k}') e^{i(k+k') \cdot x} + \omega\omega' a(\vec{k}) a^*(\vec{k}') e^{i(k-k') \cdot x} + \right. \right. \\
&\quad \left. \left. + \omega\omega' a^*(\vec{k}) a(\vec{k}') e^{-i(k-k') \cdot x} - \omega\omega' a^*(\vec{k}) a^*(\vec{k}') e^{-i(k+k') \cdot x} \right) + \right. \quad (8.18) \\
&\quad \left. + \left( -\vec{k} \cdot \vec{k}' a(\vec{k}) a(\vec{k}') e^{i(k+k') \cdot x} + \vec{k} \cdot \vec{k}' a(\vec{k}) a^*(\vec{k}') e^{i(k-k') \cdot x} + \right. \right. \\
&\quad \left. \left. + \vec{k} \cdot \vec{k}' a^*(\vec{k}) a(\vec{k}') e^{-i(k-k') \cdot x} - \vec{k} \cdot \vec{k}' a^*(\vec{k}) a^*(\vec{k}') e^{-i(k+k') \cdot x} \right) + \right. \\
&\quad \left. + m^2 \left( a(\vec{k}) a(\vec{k}') e^{i(k+k') \cdot x} + a(\vec{k}) a^*(\vec{k}') e^{i(k-k') \cdot x} + \right. \right. \\
&\quad \left. \left. + a^*(\vec{k}) a(\vec{k}') e^{-i(k-k') \cdot x} + a^*(\vec{k}) a^*(\vec{k}') e^{-i(k+k') \cdot x} \right) \right] - V\Omega,
\end{aligned}$$

where  $V$  is the volume of the space resulting from the  $\int d^3x$  integral. Then, from the fact that

$$\int d^3x e^{i\vec{x} \cdot \vec{y}} = (2\pi)^3 \delta^3(\vec{y}), \quad (8.19)$$

we have

$$\begin{aligned}
H &= \frac{1}{2} (2\pi)^3 \int \widetilde{dk} \widetilde{dk}' \left[ \delta^3(\vec{k} - \vec{k}') \left( \omega\omega' + \vec{k} \cdot \vec{k}' + m^2 \right) \times \right. \\
&\quad \times \left( a^*(\vec{k}) a(\vec{k}') e^{-i(\omega-\omega')t} + a(\vec{k}) a^*(\vec{k}') e^{-i(\omega-\omega')t} \right) + \\
&\quad + \delta^3(\vec{k} + \vec{k}') \left( -\omega\omega' - \vec{k} \cdot \vec{k}' + m^2 \right) \times \quad (8.20) \\
&\quad \times \left( a(\vec{k}) a(\vec{k}') e^{-i(\omega+\omega')t} + a^*(\vec{k}) a^*(\vec{k}') e^{i(\omega+\omega')t} \right) \left. \right] - V\Omega = \\
&= \frac{1}{2} \int \widetilde{dk} \frac{1}{2\omega} \left[ \left( \omega^2 + \vec{k}^2 + m^2 \right) \left( a^*(\vec{k}) a(\vec{k}) + a(\vec{k}) a^*(\vec{k}) \right) + \right. \\
&\quad + \left( -\omega^2 + \vec{k}^2 + m^2 \right) \left( a(\vec{k}) a(-\vec{k}) e^{-2i\omega t} + \right. \\
&\quad \left. \left. + a^*(\vec{k}) a^*(-\vec{k}) e^{2i\omega t} \right) \right] - V\Omega,
\end{aligned}$$

and finally, using the definition of  $\omega$  (equation (8.7)), this becomes

$$H = \frac{1}{2} \int \widetilde{dk} \, \omega \left( a^\star(\vec{k})a(\vec{k}) + a(\vec{k})a^\star(\vec{k}) \right) - V\Omega . \quad (8.21)$$

Using (8.17), we can rewrite this as (switching from  $\star$  to  $\dagger$  to emphasize the Hermitian nature):

$$\begin{aligned} H &= \frac{1}{2} \int \widetilde{dk} \, \omega \left[ a^\dagger(\vec{k})a(\vec{k}) + (2\pi)^3 2\omega \delta^3(\vec{k} - \vec{k}) + a^\dagger(\vec{k})a(\vec{k}) \right] - V\Omega = \\ &= \int \widetilde{dk} \, \omega a^\dagger(\vec{k})a(\vec{k}) + \int \widetilde{dk} \, \omega (2\pi)^3 \delta^3(0) - V\Omega = \\ &= \int \widetilde{dk} \, \omega a^\dagger(\vec{k})a(\vec{k}) + \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega} \omega (2\pi)^3 \delta^3(0) - V\Omega = \\ &= \int \widetilde{dk} \, \omega a^\dagger(\vec{k})a(\vec{k}) + \frac{1}{2} \delta^3(0) \int d^3\vec{k} - V\Omega . \end{aligned} \quad (8.22)$$

Notice that both the second and third terms are infinite (assuming the volume  $V$  of the space we are in is infinite). This may be troubling, but remember that  $\Omega$  is an arbitrary constant we can set to be anything we want. So, let's define

$$\Omega \equiv \frac{1}{2V} \delta^3(0) \int d^3\vec{k} , \quad (8.23)$$

leaving

$$H = \int \widetilde{dk} \, \omega a^\dagger(\vec{k})a(\vec{k}) . \quad (8.24)$$

Remember that measurement can only detect changes in energy, and therefore the infinity we subtracted off does not affect the value we will measure experimentally.

What we have done in (8.24), by subtracting off the infinite part in a way that doesn't change the physics, is a very primitive example of *Renormalization*. Often, for various reasons, measurable quantities in QFT are plagued by different types of infinities. However, it is possible to subtract off those infinities in a well-defined way, leaving a finite part. It turns out that this finite part is the correct value seen in nature. The reasons for this are very deep, and we will not discuss them (or general renormalization theory) in much depth now. For correlating theoretical results with experiment, being able to renormalize results correctly is vital. However, our goal in this section

is not to understand the subtleties of renormalization, but to understand the overall structure of particle physics.

So, we have our field expansion (8.15) and commutation relations (8.17). Notice that (8.17) have the exact form of a simple harmonic oscillator, which you learned about in introductory quantum mechanics. Therefore, because they have the same structure as the harmonic oscillator, they will have the same physics. By doing nothing but imposing relativity, we have found that scalar fields, which are Hermitian operators, act as raising and lowering (or synonymously creation and annihilation) operators on the vacuum (just like the simple harmonic oscillator).

Comparing (8.17) with the standard harmonic oscillator operators, it is clear that  $a^\dagger(\vec{k})$  creates a  $\phi$  particle with momentum  $\vec{k}$  and energy  $\omega$ , whereas  $a(\vec{k})$  annihilates a  $\phi$  particle with momentum  $\vec{k}$  and energy  $\omega$ . A normalized state will be

$$|\vec{k}\rangle = \sqrt{2\omega} a^\dagger(\vec{k})|0\rangle. \quad (8.25)$$

The entire spectrum of states can be studied by acting on  $|0\rangle$  with creation operators, and probability amplitudes for one state to be found in another,  $\langle\vec{k}_f|\vec{k}_i\rangle$ , are straightforward to calculate (and positive semi-definite). Naturally this theory does not discuss any interactions between particles, and therefore we will have to do a great deal of modification before we are done. But this simple exercise of merely imposing the standard commutation relations (8.6) between the field and its momentum, we have gained complete knowledge of the quantum mechanical states of the theory.

## 8.2 The Spin-Statistics Theorem

Notice that the states coming from (8.25) will include the two particle state

$$|\vec{k}; \vec{k}'\rangle = 2\sqrt{\omega\omega'} a^\dagger(\vec{k})a^\dagger(\vec{k}')|0\rangle. \quad (8.26)$$

But the commutation relations (8.17) tell us that  $a^\dagger(\vec{k})a^\dagger(\vec{k}') = a^\dagger(\vec{k}')a^\dagger(\vec{k})$ . So, this theory also allows the state

$$|\vec{k}'; \vec{k}\rangle = 2\sqrt{\omega'\omega} a^\dagger(\vec{k}')a^\dagger(\vec{k})|0\rangle. \quad (8.27)$$

Recall from a chemistry, or modern physics course, that particles with half-integer spin obey the Pauli Exclusion Principle, whereas particles of integer

spin do not. Our Klein-Gordon scalar fields  $\phi$  are spinless ( $j = 0$ ), and therefore we would expect that they do not obey Pauli exclusion. The fact that our commutation relations have allowed both states (8.26) and (8.27) is therefore expected. This is an indication that we quantized correctly.

But notice that this statistical result (that the scalar fields do not obey Pauli exclusion) is entirely a result of the commutation relations. Therefore, if we attempt to quantize a spin-1/2 field in the same way, they will obviously not obey Pauli exclusion either. We must therefore quantize spin-1/2 differently.

It turns out that the correct way to quantize spin-1/2 fields is to use, instead of commutation relations like we used for scalar fields, anticommutation relations. If the operators of our spin-1/2 fields obey

$$\{a_1^\dagger, a_2^\dagger\} = a_1^\dagger a_2^\dagger = 0 \quad \Rightarrow \quad a_1^\dagger a_2^\dagger = -a_2^\dagger a_1^\dagger, \quad (8.28)$$

then if we try to act twice with the same operator, we have

$$a_1^\dagger a_1^\dagger |0\rangle = -a_1^\dagger a_1^\dagger |0\rangle \quad \Rightarrow \quad a_1^\dagger a_1^\dagger |0\rangle = 0. \quad (8.29)$$

In other words, if we quantize with anticommutation relations, it is not possible for two particles to occupy the same state simultaneously.

This relationship between the spin of a particle and the statistics it obeys (which demands that integer spin particles be quantized by commutation relations and half-integer spin particles to be quantized with anticommutation relations) is called the *Spin-Statistics Theorem*.

Because particles obeying Pauli exclusion are said to have *Bose-Einstein* statistics, and particles that do not obey Pauli exclusion are said to have *Fermi-Dirac* statistics, we call particles with integer spin *Bosons*, and particles with half-integer spin *Fermions*.

### 8.2.1 Left-Handed and Right-Handed Fields

Recall that Dirac fields are the 4-component spinor  $\psi = (\psi_L, \psi_R)$ , where  $\psi_L$  transforms under the left-handed  $(0, 1/2)$  representation of the Lorentz group, and  $\psi_R$  transforms under the right-handed  $(1/2, 0)$  representation.

In general, we refer to these 2-component spinors as *Weyl* fields (see Sec. 2.4). So, the fermion is the spinor combination of two Weyl fields, one being the left-handed particle, and the other being the right-handed antiparticle.

Also we had defined the fields  $\bar{\psi} = \psi^\dagger \gamma^0 = (\psi_R^\dagger, \psi_L^\dagger)$ . If we interpret  $\bar{\psi}$  as the conjugate of  $\psi$  (which the form of the Dirac Lagrangian implies we should), then we see that the right-handed field is the conjugate of the left, and vice versa. Or, in other words,

$$\psi_L^\dagger = \psi_R \quad \text{and} \quad \psi_R^\dagger = \psi_L . \quad (8.30)$$

We take advantage of the fact by writing all fields in terms of left-handed Weyl fields. For example, given the two left-handed Weyl fields  $\chi$  and  $\xi$ , we can form the 4-component spinor field  $\psi = (\chi, \xi^\dagger)$ , and so  $\bar{\psi} = (\xi, \chi^\dagger)$ . We will refer to such a field as a *Dirac Field*, and denote it  $\psi_D$  (see Sec. 2.5).

On the other hand, we could define a 4-component spinor in terms of a single left-handed Weyl field  $\chi$ , or  $\psi = (\chi, \chi^\dagger)$ . But now notice that  $\bar{\psi} = (\chi, \chi^\dagger)$ , which is equal simply to the transpose of  $\psi$ . We refer to such a field (whose conjugate is equal to its transpose) as a *Majorana Field*, and denote it  $\psi_M$  (see Sec. 2.5).

Recall that an antiparticle has the same mass but opposite charge and opposite handedness of its particle. So, working with the Dirac field  $\psi_D$ , we can change the charge by merely swapping  $\chi$  and  $\xi$ , using the *Charge Conjugation* operator  $\mathcal{C}$  defined by

$$\mathcal{C}\psi_D = \mathcal{C} \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix} = \begin{pmatrix} \xi \\ \chi^\dagger \end{pmatrix} . \quad (8.31)$$

Also, consider the transpose of  $\bar{\psi}_D$  (which is just returning the conjugate of  $\psi_D$  to column form). Acting on this with  $\mathcal{C}$  gives

$$\mathcal{C}\bar{\psi}_D^T = \mathcal{C} \begin{pmatrix} \xi \\ \chi^\dagger \end{pmatrix} = \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix} = \psi_D . \quad (8.32)$$

So, we have

$$\mathcal{C}\psi_D = \bar{\psi}_D^T \quad \text{and} \quad \mathcal{C}\bar{\psi}_D^T = \psi_D . \quad (8.33)$$

We therefore say that  $\psi_D$  and  $\bar{\psi}_D^T$  are *Charge Conjugate* to each other.

However, notice that with the Majorana field,

$$\mathcal{C}\psi_M = \mathcal{C} \begin{pmatrix} \chi \\ \chi^\dagger \end{pmatrix} = \begin{pmatrix} \chi \\ \chi^\dagger \end{pmatrix} = \psi_M , \quad (8.34)$$

and

$$\mathcal{C}\bar{\psi}_M^T = \bar{\psi}_M^T = \psi_M . \quad (8.35)$$

In summary, Dirac fields are not equal to their charge conjugate, while Majorana fields are. By analogy with scalars (where the complex conjugate of a real number is equal to itself, whereas the complex conjugate of a complex number is not), we often refer to Majorana fields as *Real*, and to Dirac fields as *Complex*.

So, we can now write out the Lagrangian for Dirac and Majorana fields in terms of their Weyl fields:

$$\mathcal{L}_D = i\chi^\dagger \bar{\sigma}^\mu \partial_\mu \chi + i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \xi - m (\chi\xi + \chi^\dagger \xi^\dagger) , \quad (8.36)$$

$$\mathcal{L}_M = i\chi^\dagger \bar{\sigma}^\mu \partial_\mu \chi - \frac{1}{2}m (\chi\chi + \chi^\dagger \chi^\dagger) . \quad (8.37)$$

### 8.3 Canonical Quantization of Fermions

The general solution to the Dirac equation is

$$\psi_D(x) = \sum_{s=1}^2 \int \widetilde{d\vec{k}} [b_s(\vec{k})u_s(\vec{k})e^{ik\cdot x} + d_s^\dagger(\vec{k})v_s(\vec{k})e^{-ik\cdot x}] , \quad (8.38)$$

where  $s=1, 2$  are the two spin states,  $b_s$  and  $d_s^\dagger$  are (respectively) the lowering operator for the particle and the raising operator for the antiparticle. The charge conjugate of  $\psi_D$  will have the raising operator for the particle and the lowering operator for the antiparticle. The  $u_s$  and  $v_s$  are constant 4-component vectors which act as a basis for all particle/antiparticle states in the spinor space (for our purposes, they are merely present to make  $\psi_D$  a 4-component field).

We quantize, as we said in section 8.2, using anti-commutation relations. Writing only the non-zero relation,

$$\{\psi_\alpha(t, \vec{x}), \bar{\psi}_\beta(t, \vec{x}')\} = \delta^3(\vec{x} - \vec{x}')(\gamma^0)_{\alpha\beta} . \quad (8.39)$$

These imply that the only non-zero commutation relations in terms of the operators are

$$\begin{aligned} \{b_s(\vec{k}), b_{s'}^\dagger(\vec{k}')\} &= (2\pi)^3 \delta^3(\vec{k} - \vec{k}') 2\omega \delta_{ss'} , \\ \{d_s^\dagger(\vec{k}), d_{s'}(\vec{k}')\} &= (2\pi)^3 \delta^3(\vec{k} - \vec{k}') 2\omega \delta_{ss'} . \end{aligned} \quad (8.40)$$



These form the algebra of a simple harmonic oscillator, and we can therefore find the entire spectrum of states by acting on  $|0\rangle$  with  $b_s^\dagger$  and  $d_s^\dagger$ .

Then, following a series of calculations nearly identical to the ones in section 8.1, we arrive at the Hamiltonian

$$H = \sum_{s=1}^2 \int \widetilde{d\vec{k}} \omega [b_s^\dagger(\vec{k})b_s(\vec{k}) + d_s^\dagger(\vec{k})d_s(\vec{k})] - \lambda, \quad (8.41)$$

where  $\lambda$  is an infinite constant we can merely subtract off and therefore ignore.

Comparing (8.24) and (8.41), we see that they both have essentially the same form;  $\omega$  (which is energy) to the left of the creation operator, which is to the left of the annihilation operator. To understand the meaning of this, we will see how it generates energy eigenvalues. We will use equation (8.24) for simplicity. Consider acting with the Hamiltonian operator on some arbitrary state  $|\vec{p}\rangle$  with momentum  $\vec{p}$ . Using (8.25),

$$\begin{aligned} H|\vec{p}\rangle &= \int \widetilde{d\vec{k}} \omega_k a^\dagger(\vec{k})a(\vec{k})|\vec{p}\rangle = \int \widetilde{d\vec{k}} \omega_k a^\dagger(\vec{k})a(\vec{k})\sqrt{2\omega_p}a^\dagger(\vec{p})|0\rangle = \\ &= \int \widetilde{d\vec{k}} \omega_k \sqrt{2\omega_p}a^\dagger(\vec{k}) \left( (2\pi)^3 2\omega_p \delta^3(\vec{k} - \vec{p}) + a^\dagger(\vec{p})a(\vec{k}) \right) |0\rangle = \\ &= \int \widetilde{d\vec{k}} \omega_k \sqrt{2\omega_p}a^\dagger(\vec{k}) (2\pi)^3 2\omega_p \delta^3(\vec{k} - \vec{p}) |0\rangle = \quad (8.42) \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3 2\omega_k} \omega_k \sqrt{2\omega_p}a^\dagger(\vec{k}) (2\pi)^3 2\omega_p \delta^3(\vec{k} - \vec{p}) |0\rangle = \\ &= \int d^3\vec{k} \sqrt{2\omega_p}a^\dagger(\vec{k}) \omega_p \delta^3(\vec{k} - \vec{p}) |0\rangle = \\ &= \omega_p \sqrt{2\omega_p}a^\dagger|0\rangle = \omega_p |\vec{p}\rangle. \end{aligned}$$

So,  $H|\vec{p}\rangle = \omega_p |\vec{p}\rangle$ , where  $\omega_p = \vec{p}^2 + m^2$ , which is the relativistic equation for energy as in equation (8.7). So, the Hamiltonian operator gives the appropriate energy eigenvalue on our physical quantum states.

For the Dirac Hamiltonian the eigenvalue will be a linear combination of the energies of each type of particle. If we denote the states as  $|\vec{p}_b, s_b; \vec{p}_d, s_d\rangle$ , where the first two elements give the state of a  $b$  type particle and the second of the  $d$  type particle, we have

$$H|\vec{p}_b, s_b; \vec{p}_d, s_d\rangle = \dots = (\omega_{p_b} + \omega_{p_d})|\vec{p}_b, s_b; \vec{p}_d, s_d\rangle. \quad (8.43)$$

For Majorana fields we only have one type of particle,

$$\psi_M(x) = \sum_{s=1}^2 \int \widetilde{d\vec{k}} \left[ b_s(\vec{k}) u_s(\vec{k}) e^{ik \cdot x} + b_s^\dagger(\vec{k}) v_s(\vec{k}) e^{-ik \cdot x} \right], \quad (8.44)$$

and quantization with anticommutation relations will give

$$H = \sum_{s=1}^2 \int \widetilde{d\vec{k}} \omega b_s^\dagger(\vec{k}) b_s(\vec{k}). \quad (8.45)$$

**Exercise 8.1:** Find vacuum expectation value of the scalar field Hamiltonian.

**Exercise 8.2:** Prove that the parity operator of a scalar field commutes with its Hamiltonian.

**Exercise 8.3:** Suppose to the Lagrangian for the real scalar field,  $\phi(x)$ , one adds the term  $J(x)\phi(x)$ , where  $J(x)$  is a real c-number function. Calculate  $\langle 0|\phi(x)|0\rangle$  and the two-point function  $\langle 0|T(\phi(x)\phi(0))|0\rangle$  exactly. Treat the term  $J(x)\phi(x)$  as a perturbation and calculate these quantities to the lowest order in  $J(x)$ .

**Exercise 8.4:** If the Dirac field is quantized according to the Bose-Einstein rather than Fermi-Dirac statistics, what would be the energy of the field?

**Exercise 8.5:** Express the Majorana field operator using creation and annihilation operators of a Dirac field. Introduce creation and annihilation operators for Majorana spinors and find corresponding anticommutation relations. Rewrite the QED Lagrangian density using Majorana spinors.

**Exercise 8.6:** Calculate the commutators between components of the electric and the magnetic fields.

**Exercise 8.7:** Express the angular momentum of the photon field in terms of the potentials in the Coulomb gauge.

**Exercise 8.8:** Quantize the electromagnetic field between two parallel, perfect conductor square plates.

## Chapter 9

# Path Integral Quantization

While the *Canonical Quantization* procedure giving us a tremendous amount of information (the entire spectrum of states for particles), it is still lacking quite a bit. As we said at the beginning of Sec. 7.1, we ultimately want a relativistic quantum mechanical theory of interactions. Canonical Quantization has provided a relativistic quantum mechanical theory, but we aren't close to being able to incorporate interactions into our theory. While it is possible to incorporate interactions, it is very difficult, and in order to simplify we will need a new way of quantizing.

Perhaps the most fundamental experiment in quantum mechanics is the *Double Slit* experiment. In brief, what this experiment tells us is that, when a single electron moves through a screen with two slits, and no observation is made regarding which slit it goes through, it actually goes through both slits, and until a measurement is made (for example, when it hits the observation screen behind the double slit), it exists in a superposition of both paths.

As a result, the particle exhibits a wave nature, and the pattern that emerges on the observation screen is an interference pattern – the same as if a classical wave was passing through the double slit—all paths in the superposition of the single electron are interfering with each other, both destructively and constructively. Once the electron is observed on the observation screen, it collapses probabilistically into one of its possible states (a particular location on the observation screen).

If, on the other hand, you set up some mechanism to observe which of the

two slits the electron travels through, then the observation has been made before the observation screen, and you no longer have the superposition, and therefore you no longer see any indication of an interference pattern. The electrons are behaving, in a sense, classically from the double slit to the observation screen in this case.

The meaning of this is that a particle that has not been observed will actually take every possible path at once. Once an observation has been made, there is some probability associated with each path. Some paths are very likely, and others are less likely (some are nearly impossible). But until observation, it actually exists in a superposition of all possible states/paths.

So, to quantize, we will create a mathematical expression for a 'sum over all possible paths'. This expression is called a *Path Integral*, and will prove to be a much more useful way to quantize a physical system.

We begin this construction by considering merely the amplitude for a particle at position  $q_1$  at time  $t_1$  to propagate to  $q_2$  at time  $t_2$ . This amplitude will be given by

$$\langle q_2, t_2 | q_1, t_1 \rangle = \langle q_2 | e^{iH(t_2-t_1)} | q_1 \rangle . \quad (9.1)$$

To evaluate this, we begin by dividing the time interval  $t_2 - t_1$  into  $N + 1$  equal intervals of length

$$\delta t = \frac{t_2 - t_1}{N + 1} \quad (9.2)$$

each. So, we can insert  $N$  complete sets of position eigenstates,

$$\begin{aligned} \langle q_2, t_2 | q_1, t_1 \rangle = \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \langle q_2 | e^{-iH\delta t} | Q_N \rangle \langle Q_N | e^{-iH\delta t} | Q_{N-1} \rangle \cdots \\ \cdots \langle Q_1 | e^{-iH\delta t} | q_1 \rangle . \end{aligned} \quad (9.3)$$

Let's look at a single one of these amplitudes. We know that in nearly all physical theories, we can break the Hamiltonian up as

$$H = \frac{P^2}{2m} + V(Q) . \quad (9.4)$$

So, using the completeness of momentum eigenstates,

$$\begin{aligned}
\langle Q_{i+1}|e^{-iH\delta t}|Q_i\rangle &= \langle Q_{i+1}|e^{-i(P^2/2m+V(Q))\delta t}|Q_i\rangle = \\
&= \langle Q_{i+1}|e^{-i\delta t P^2/2m}e^{-i\delta t V(Q)}|Q_i\rangle = \\
&= \int dP' \langle Q_{i+1}|e^{-i\delta t P^2/2m}|P'\rangle \langle P'|e^{-i\delta t V(Q)}|Q_i\rangle = \\
&= \int dP' e^{-i\delta t P'^2/2m} e^{-i\delta t V(Q_i)} \langle Q_{i+1}|P'\rangle \langle P'|Q_i\rangle = \\
&= \int dP' e^{-i\delta t P'^2/2m} e^{-i\delta t V(Q_i)} \frac{e^{iP'Q_{i+1}}}{\sqrt{2\pi}} \frac{e^{-iP'Q_i}}{\sqrt{2\pi}} = \\
&= \int \frac{dP'}{2\pi} e^{iH\delta t} e^{iP'(Q_{i+1}-Q_i)} = \tag{9.5} \\
&= \int \frac{dP'}{2\pi} e^{i[P'(Q_{i+1}-Q_i)-H\delta t]} = \\
&= \int \frac{dP'}{2\pi} e^{i\delta t [P'(Q_{i+1}-Q_i)/\delta t - H]} .
\end{aligned}$$

And taking the limit as  $\delta t \rightarrow 0$ ,  $(Q_{i+1} - Q_i)/\delta t \rightarrow \dot{Q}_i$ . So,

$$\int \frac{dP'}{2\pi} e^{i\delta t [P'(\frac{Q_{i+1}-Q_i}{\delta t}) - H]} = \int \frac{dP'}{2\pi} e^{idt_{i+1}[P'\dot{Q}_i - H]} , \tag{9.6}$$

where the subscript on  $dt$  merely indicates where the infinitesimal time interval 'ends'. So, we can plug this into (9.3) and taking the limit as  $\delta t \rightarrow 0$ ,

$$\begin{aligned}
\langle q_2, t_2 | q_1, t_1 \rangle &= \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \langle q_2 | e^{-iH\delta t} | Q_N \rangle \langle Q_N | e^{-iH\delta t} | Q_{N-1} \rangle \cdots \\
&\cdots \langle Q_1 | e^{-iH\delta t} | q_1 \rangle = \lim_{N \rightarrow \infty} \int_{-\infty}^{\infty} \prod_{i=1}^N dQ_i \int \frac{dP'_i}{2\pi} \times \tag{9.7} \\
&\times e^{idt_2 [P'_N \dot{Q}_N - H]} e^{idt_N [P'_{N-1} \dot{Q}_{N-1} - H]} \cdots e^{idt_1 [P'_1 \dot{Q}_1 - H]} = \\
&= \int_{-\infty}^{\infty} \mathcal{D}p \mathcal{D}q e^{-i \int_{t_1}^{t_2} dt (pq - H)} ,
\end{aligned}$$

where

$$\mathcal{D}p = \prod_{i=1}^{\infty} dp_i , \quad \mathcal{D}q = \prod_{i=1}^{\infty} dq_i . \tag{9.8}$$

If  $p$  shows up quadratically (as it always does in  $p^2/2m$ ), or linearly (e.g. flat wave phase,  $ipq$ ), then we can merely do the Gaussian integral over  $p$ ,

using the Gaussian integral formula:

$$\int_{-\infty}^{\infty} dx e^{(-ax^2+bx)} = \sqrt{\frac{\pi}{a}} e^{b^2/4a} , \quad (9.9)$$

where  $a$  and  $b$  are some constants. This will result in an overall constant which we merely absorb back into the measure when we normalize. Then, recognizing that the integrand in the exponent is  $p\dot{q} - H = \mathcal{L}$ , we have

$$\langle q_2, t_2 | q_1, t_1 \rangle = \int \mathcal{D}q e^{i \int_{t_1}^{t_2} dt \mathcal{L}} = \int \mathcal{D}q e^{iS} . \quad (9.10)$$

Formally, the measure of (9.10) has an infinite number of differentials, and therefore evaluating it would require doing an infinite number of integrals. This is to be expected, since the point of the path integral is a sum over every possible path, of which there are an infinite number. So, because we obviously can't do an infinite number of integrals, we will have to find a clever way of evaluating (9.10). But before doing so, we discuss what the path integral means.

## 9.1 Interpretation of the Path Integral

Equation (9.10) says that, given an initial and final configuration  $(q_1, t_1)$  and  $(q_2, t_2)$ , absolutely any path between them is possible. This is the content of the  $\mathcal{D}q$  part: it is the sum over all paths. Then, for each of those paths, the integral assigns a *statistical weight* of  $e^{iS}$  to it, where the action  $S$  is calculated using that path (recall our comments in section 1.1 about  $S$  being a functional, not a function).

So, consider an arbitrary path  $q_0$ , which receives statistical weight  $e^{iS[q_0]}$ . Now, consider a path  $q'$  very close to  $q_0$ , only varying by a small amount:  $q' = q_0 + \epsilon \delta q_0$ . This will have statistical weight

$$e^{iS[q_0 + \epsilon \delta q_0]} = e^{iS[q_0] + i\epsilon \delta q_0 \delta S[q_0] / \delta q} , \quad (9.11)$$

where  $\delta S / \delta q$  is the Euler Lagrange derivative (1.6)

$$\frac{\delta S}{\delta q} = \frac{d}{dt} \frac{\partial S}{\partial \dot{q}} - \frac{\partial S}{\partial q} . \quad (9.12)$$

To make our intended result more obvious, we do a Wick rotation, taking  $t \rightarrow it$ , so  $dt \rightarrow idt$ , and

$$S = \int dt \mathcal{L} \quad \rightarrow \quad i \int dt \mathcal{L} = iS, \quad (9.13)$$

and  $e^{iS} \rightarrow e^{-S}$ . Now, the path  $q' = q_0 + \epsilon \delta q_0$  gets weight

$$e^{S[q_0]} e^{-\epsilon \delta q_0 \delta S[q_0] / \delta q}. \quad (9.14)$$

If  $\delta S / \delta q$  is very large, then the weight becomes exponentially small, i.e. the larger the variation of the action is, the less probable that path is.

So the most probable path is the one for the smallest value of  $\delta S / \delta t$ , or the path at which  $\delta S / \delta q = 0$ . And as we discussed in 1.1, this is the path of *Least Action*. Thus, we have recovered classical mechanics as the first order approximation of quantum mechanics.

So, the meaning of the path integral is that all imaginable paths are possible for the particle to travel in moving from one configuration to another. However, not all paths are equally probable. The likelihood of a given path is given by the action exponentiated, and therefore the most probable paths are the ones which minimize the action. This is the reason that, macroscopically, the world appears classical. The likelihood of every particle in, say, a baseball, simultaneously taking a path noticeably far from the path of least action is negligibly small.

We will find that path integral quantization provides an extremely powerful tool with which to create our relativistic quantum theory of interactions.

## 9.2 Expectation Values

Now that we have a way of finding  $\langle q_2, t_2 | q_1, t_1 \rangle$ , the natural question to ask next is how do we find expectation values like  $\langle q_2, t_2 | Q(t') | q_1, t_1 \rangle$  or  $\langle q_2, t_2 | P(t') | q_1, t_1 \rangle$ . By doing a similar derivation as above, it is easy to show that

$$\langle q_2, t_2 | Q(t') | q_1, t_1 \rangle = \dots = \int \mathcal{D}q Q(t') e^{iS}. \quad (9.15)$$

We will find that evaluating integrals of this form is simplified greatly through making use of *Functional Derivatives*. For some function  $f(x)$ , the functional

derivative is defined by

$$\frac{\delta}{\delta f(y)} f(x) \equiv \delta(x - y) . \quad (9.16)$$

Next, we modify our path integral by adding an *Auxiliary External Source* function, so that

$$\mathcal{L} \rightarrow \mathcal{L} + f(t)Q(t) + h(t)P(t) . \quad (9.17)$$

So we now have

$$\langle q_2, t_2 | q_1, t_1 \rangle_{f,h} = \int \mathcal{D}q e^{\int dt (\mathcal{L} + fQ + hP)} , \quad (9.18)$$

which allows us to write out expectation values in the simple form

$$\begin{aligned} \langle q_2, t_2 | Q(t') | q_1, t_1 \rangle &= \frac{1}{i} \frac{\delta}{\delta f(t')} \langle q_2, t_2 | q_1, t_1 \rangle_{f,h} \Big|_{f,h=0} = \\ &= \int \mathcal{D}q Q(t') e^{iS + i \int dt (fQ + hP)} \Big|_{f,h=0} = \\ &= \int \mathcal{D}q Q(t') e^{iS} , \end{aligned} \quad (9.19)$$

or

$$\begin{aligned} \langle q_2, t_2 | P(t') | q_1, t_1 \rangle &= \frac{1}{i} \frac{\delta}{\delta h(t')} \langle q_2, t_2 | q_1, t_1 \rangle_{f,h} \Big|_{f,h=0} = \\ &= \int \mathcal{D}q P(t') e^{iS + i \int dt (fQ + hP)} \Big|_{f,h=0} = \\ &= \int \mathcal{D}q P(t') e^{iS} . \end{aligned} \quad (9.20)$$

So, once we have  $\langle q_2, t_2 | q_1, t_1 \rangle$ , we can find any expectation value we want simply by taking successive functional derivatives.

### 9.3 Path Integrals with Fields

Because we can build whatever state we want by acting on the vacuum, the important quantity for us to work with will be the *Vacuum to Vacuum* expectation value, or VEV,  $\langle 0|0 \rangle$ , and the various expectation values we can build through functional derivatives ( $\langle 0|\phi\phi|0 \rangle$ ,  $\langle 0|\psi\phi\phi|0 \rangle$ , etc.).



For simplicity let's write the path integral a scalar boson  $\phi$ ,

$$\langle 0|0\rangle = \int \mathcal{D}\phi e^{i \int d^4x [-\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2]} \equiv \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_0} . \quad (9.21)$$

We will eventually want to find expectation values, so we introduce the auxiliary field  $J$ , creating *the Generating Functional*:

$$Z_0(J) \equiv \langle 0|0\rangle_J = \int \mathcal{D}\phi e^{i \int d^4x (\mathcal{L}_0 + J\phi)} . \quad (9.22)$$

So, for example,

$$\langle 0|\phi|0\rangle = \frac{1}{i} \frac{\delta}{\delta J} \langle 0|0\rangle_J \Big|_{J=0} . \quad (9.23)$$

Of course, we still have a path integral with an infinite number of integrals to evaluate. But, we are finally able to discuss how we can do the evaluation.

Making use of the Fourier Transform of  $\phi$ ,

$$\tilde{\phi}(k) = \int d^4x e^{-ikx} \phi(x) , \quad \phi(x) = \int \frac{d^4k}{(2\pi)^4} e^{ikx} \tilde{\phi}(k) , \quad (9.24)$$

we begin with the  $\mathcal{L}_0$  part:

$$\begin{aligned} S_0 &= \int d^4x \mathcal{L}_0 = \int d^4x \left( -\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2 \right) = \\ &= \int d^4x \left[ -\frac{1}{2}\partial^\mu \left( \int \frac{d^4k}{(2\pi)^4} e^{ik\cdot x} \tilde{\phi}(k) \right) \partial_\mu \left( \int \frac{d^4k'}{(2\pi)^4} e^{ik'\cdot x} \tilde{\phi}(k') \right) - \right. \\ &\quad \left. -\frac{1}{2}m^2 \left( \int \frac{d^4k}{(2\pi)^4} e^{ik\cdot x} \tilde{\phi}(k) \right) \left( \int \frac{d^4k'}{(2\pi)^4} e^{ik'\cdot x} \tilde{\phi}(k') \right) \right] = \\ &= \int d^4x \left[ \frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} e^{ik\cdot x} e^{ik'\cdot x} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) \right] = \quad (9.25) \\ &= \frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) \int d^4x e^{i(k+k')\cdot x} = \\ &= \frac{1}{2} \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{\phi}(k) \tilde{\phi}(k') (k^\mu k'_\mu - m^2) (2\pi)^4 \delta^4(k+k') = \\ &= -\frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \tilde{\phi}(k) (k^2 + m^2) \tilde{\phi}(-k) . \end{aligned}$$

Then, transforming the auxiliary field part,

$$\begin{aligned}
\int d^4x J(x)\phi(x) &= \int d^4x \left( \int \frac{d^4k}{(2\pi)^4} e^{ik \cdot x} \tilde{J}(k) \right) \left( \int \frac{d^4k'}{(2\pi)^4} e^{ik' \cdot x} \tilde{\phi}(k') \right) = \\
&= \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{J}(k) \tilde{\phi}(k') \int d^4x e^{i(k+k') \cdot x} = \\
&= \int \frac{d^4k d^4k'}{(2\pi)^8} \tilde{J}(k) \tilde{\phi}(k') (2\pi)^4 \delta^4(k+k') = \tag{9.26} \\
&= \int \frac{d^4k}{(2\pi)^4} \tilde{J}(k) \tilde{\phi}(-k) .
\end{aligned}$$

Because the integral is over all  $k^\mu$ , we can rewrite this as

$$\int \frac{d^4k}{(2\pi)^4} \tilde{J}(k) \tilde{\phi}(-k) = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left( \tilde{J}(k) \tilde{\phi}(-k) + \tilde{J}(-k) \tilde{\phi}(k) \right) \tag{9.27}$$

(we did this to get the factor of 1/2 out front in order to have the same coefficient as the  $\mathcal{L}_0$  part from above).

So,

$$\begin{aligned}
S = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left[ -\tilde{\phi}(k)(k^2 + m^2)\tilde{\phi}(-k) + \right. \\
\left. + \tilde{J}(k)\tilde{\phi}(-k) + \tilde{J}(-k)\tilde{\phi}(k) \right] . \tag{9.28}
\end{aligned}$$

Now, we make a change of variables,

$$\tilde{\chi}(k) \equiv \tilde{\phi}(k) - \frac{\tilde{J}(k)}{k^2 + m^2} \tag{9.29}$$

(Note that this leaves the measure of the path integral unchanged:  $\mathcal{D}\phi \rightarrow \mathcal{D}\chi$ ). Plugging this we have,

$$\begin{aligned}
S &= \int \frac{d^4k}{2(2\pi)^4} \left[ -\left( \tilde{\chi}(k) + \frac{\tilde{J}(k)}{k^2 + m^2} \right) (k^2 + m^2) \left( \tilde{\chi}(-k) + \frac{\tilde{J}(-k)}{k^2 + m^2} \right) + \right. \\
&+ \tilde{J}(k) \left( \tilde{\chi}(-k) + \frac{\tilde{J}(-k)}{k^2 + m^2} \right) + \tilde{J}(-k) \left( \tilde{\chi}(k) + \frac{\tilde{J}(k)}{k^2 + m^2} \right) \left. \right] = \tag{9.30} \\
&= \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \left[ -\tilde{\chi}(k)(k^2 + m^2)\tilde{\chi}(-k) + \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2 + m^2} \right]
\end{aligned}$$

(The point of all of this is that, in this form, we have all of the  $\phi$ , or equivalently  $\chi$ , dependence in the first term, with no  $\phi$  or  $\chi$  dependence on the second term). Finally, our generating functional (9.22) is

$$\langle 0|0\rangle_J = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4k}{(2\pi)^4} \left[ -\tilde{\chi}(k)(k^2+m^2)\tilde{\chi}(-k) + \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2+m^2} \right]} . \quad (9.31)$$

Using some physical reasoning, we can see how to evaluate the infinite number of integrals in this expression. Notice that if we set  $J = 0$ , we have a free theory in which no interactions take place. This means that if we start with nothing (the vacuum), the probability of having nothing later is 100%,

$$\langle 0|0\rangle_J|_{J=0} = 1 = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4k}{(2\pi)^4} \left[ -\tilde{\chi}(k)(k^2+m^2)\tilde{\chi}(-k) \right]} . \quad (9.32)$$

If that part is 1, then we have

$$\langle 0|0\rangle_J = \int \mathcal{D}\chi e^{\frac{i}{2} \int \frac{d^4k}{(2\pi)^4} \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2+m^2}} . \quad (9.33)$$

Remarkably, the integrand has no  $\chi$  dependence! Therefore, the infinite number of integrals over all possible paths becomes nothing more than a constant we can absorb into the normalization, leaving

$$\langle 0|0\rangle_J = e^{\frac{i}{2} \int \frac{d^4k}{(2\pi)^4} \frac{\tilde{J}(k)\tilde{J}(-k)}{k^2+m^2}} . \quad (9.34)$$

We can Fourier Transform back to coordinate space to get

$$Z_0(J) = \langle 0|0\rangle_J = e^{\frac{i}{2} \int d^4x d^4x' J(x)\Delta(x-x')J(x')} , \quad (9.35)$$

where

$$\Delta(x-x') \equiv \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik \cdot (x-x')}}{k^2+m^2} \quad (9.36)$$

is called the *Feynman Propagator* for the scalar field. We can then find expectation values by operating on this with  $\delta/i\delta J$  as described in Sec. 9.2.

We can repeat everything we have just done for fermions, and while it is a great deal more complicated (and tedious), it is in essence the same calculation. We begin by adding the auxiliary function  $\bar{\eta}\psi + \bar{\psi}\eta$ , to get expectation values of  $\bar{\psi}$  and  $\psi$  by using  $\delta/i\delta\eta$  and  $\delta/i\delta\bar{\eta}$ , respectively.

We then Fourier Transform every term in the exponent and find that we can separate out the  $\bar{\psi}$  and  $\psi$  dependence, allowing us to set the term which does depend on  $\psi$  and  $\bar{\psi}$  equal to 1. Fourier Transforming back then gives

$$Z_0(\eta, \bar{\eta}) = e^{i \int d^4x d^4x' \bar{\eta}(x) S(x-x') \eta(x')} , \quad (9.37)$$

where

$$S(x-x') = \int \frac{d^4k}{(2\pi)^4} \frac{(-\gamma^\mu k_\mu + m)}{k^2 + m^2} e^{ik \cdot (x-x')} \quad (9.38)$$

is the Feynman propagator for fermion fields.

Recall that we are calling the auxiliary fields  $J$ ,  $\eta$ , and  $\bar{\eta}$  *Source Fields*. Comparing the form of the Lagrangian in equation (9.22) to (2.6) reveals why.  $J$ ,  $\eta$ , and  $\bar{\eta}$  behave mathematically as sources, giving rise to the field they are coupled to, in the same way that the electromagnetic source  $J^\mu$  gives rise to the electromagnetic field  $A^\mu$ . The meaning behind equations (9.35) (and (9.37)) is that  $J$  (or  $\eta$  and  $\bar{\eta}$ ) act as sources for the fields, creating a  $\phi$  (or  $\psi$  and  $\bar{\psi}$ ) at spacetime point  $x$ , and absorbing it at point  $x'$ . The terms  $\Delta(x-x')$  and  $S(x-x')$  then represent the expression giving the probability amplitude  $\langle 0|0 \rangle$  for that particular event to occur. In other words, the propagator is the statistical weight of a particle going from  $x$  to  $x'$ .

## 9.4 Interacting Scalars and Feynman Diagrams

We can now consider how to incorporate interactions into our formalism, allowing us to finally have our relativistic quantum theory of interactions.

Beginning with the free scalar Lagrangian we can add an interaction term  $\mathcal{L}_1$ . At this point, we only have one type of particle,  $\phi$ , so we can only have  $\phi$ 's interacting with other  $\phi$ 's. Terms proportional to  $\phi$  or  $\phi^2$  are either constant or linear in the equations of motion, and therefore aren't valid candidates for interaction terms. So, the simplest expression we can have is

$$\mathcal{L}_1 = \frac{1}{3!} g \phi^3 , \quad (9.39)$$

where  $1/3!$  is a conventional normalization, and  $g$  is a *Coupling Constant*. So our total Lagrangian is

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 = -\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 + \frac{1}{6} g \phi^3 , \quad (9.40)$$

and the path integral is

$$\begin{aligned}
 Z(J) &= \langle 0|0 \rangle_J = \int \mathcal{D}\phi e^{i \int d^4x [\mathcal{L}_0 + \mathcal{L}_1 + J\phi]} = \\
 &= \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_1} e^{i \int d^4x [\mathcal{L}_0 + J\phi]} = \\
 &= \int \mathcal{D}\phi e^{i \int d^4x \mathcal{L}_1} Z_0(J) .
 \end{aligned} \tag{9.41}$$

But, recall that we can bring out a factor of  $\phi$  from  $\langle 0|0 \rangle_J$  using the functional derivative  $\delta/i\delta J$ . So, we can make the replacement

$$\mathcal{L}_1(\phi) \rightarrow \mathcal{L}_1\left(\frac{1}{i} \frac{\delta}{\delta J}\right) \Rightarrow \frac{1}{6} g \phi^3 \rightarrow \frac{g}{6} \left(\frac{1}{i} \frac{\delta}{\delta J}\right)^3 . \tag{9.42}$$

Notice that once this is done, there is no longer any  $\phi$  dependence in  $Z(J)$ . So, with the free theory, we were able to remove the  $\phi$  dependence, leading to (9.35). Here, we were able to remove it from the interaction term as well. Once again, the infinite number of integrals in (9.10) will merely give a constant which we can absorb into the normalization. This leaves the result

$$\begin{aligned}
 Z(J) &= e^{\frac{i}{6} g \int d^4x \left(\frac{1}{i} \frac{\delta}{\delta J(x)}\right)^3} Z_0(J) = \\
 &= e^{-\frac{1}{6} g \int d^4x \left(\frac{\delta}{\delta J(x)}\right)^3} e^{\frac{i}{2} \int d^4x d^4x' J(x) \Delta(x-x') J(x')} .
 \end{aligned} \tag{9.43}$$

Now, we can do two separate Taylor expansions to these two exponentials,

$$\begin{aligned}
 Z(J) &= \sum_{V=0}^{\infty} \frac{1}{V!} \left[ -\frac{g}{6} \int d^4x \left(\frac{\delta}{\delta J(x)}\right)^3 \right]^V \times \\
 &\quad \times \sum_{P=0}^{\infty} \frac{1}{P!} \left[ \frac{i}{2} \int d^4y d^4z J(y) \Delta(y-z) J(z) \right]^P .
 \end{aligned} \tag{9.44}$$

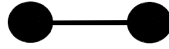
Recall that a functional derivative  $\delta/i\delta J$ , will remove a  $J$  term. Furthermore, after taking the functional derivatives, we will set  $J = 0$  to get the physical result. For a term to survive, the  $2P$  sources must all be exactly removed by the  $3V$  functional derivatives. Using (9.44), we can expand in orders of  $g$  (the coupling constant), keeping only the terms which survive, and after removing the sources, evaluate the integrals over the propagators  $\Delta$ . Then the value of the integral will be the physical amplitude for a particular event.

In practice, a slightly different formalism is used to organize and keep track of each term in this expansion. Note that there will be  $P$  propagators  $\Delta$ . We can represent each of these terms diagrammatically, by making each source a solid dot, each propagator a line, and let the  $g$  terms be vertices joining the lines together. There will be a total of  $V$  vertices, each joining 3 lines (matching the fact that we are looking at  $\phi^3$  theory; there would be 4 lines at each vertex for  $\phi^4$  theory, etc.).

For example, for  $V = 0$  and  $P = 1$ ,

$$Z(J) = \frac{i}{2} \int d^4y d^4z J(y) \Delta(y-z) J(z) . \quad (9.45)$$

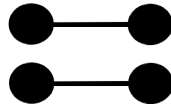
We have two sources, one located at  $z$  and the other at  $y$ , so we draw two dots, corresponding those locations. Then, the propagator  $\Delta(y-z)$  connects them together, so we draw a line between the two dots, i.e. the diagram is:



As another example, consider  $V = 0$  and  $P = 2$ . Now,

$$Z(J) = \frac{1}{2!} \left(\frac{i}{2}\right)^2 \int d^4y d^4z d^4y' d^4z' \times \\ \times [J(y) \Delta(y-z) J(z)] [J(y') \Delta(y'-z') J(z')] . \quad (9.46)$$

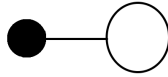
This corresponds to 4 sources at  $y, z, y'$  and  $z'$ , with propagator lines connecting  $y$  to  $z$ , and  $y'$  to  $z'$ . But, there are no lines connecting an unprimed source to a primed source, so this results in two disconnected diagrams:



As another example, consider  $V = 1$  and  $P = 2$ ,

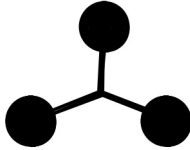
$$\begin{aligned}
 Z(J) &= -\frac{g}{6} \int d^4x \left( \frac{\delta}{\delta J(x)} \right)^3 \frac{1}{2!} \left( \frac{i}{2} \right)^2 \times \\
 &\times \int d^4y d^4z d^4y' d^4z' [J(y)\Delta(y-z)J(z)] [J(y')\Delta(y'-z')J(z')] = \\
 &= \frac{g}{48} \int d^4x d^4y d^4z d^4y' d^4z' \times \tag{9.47} \\
 &\times \delta(y-x)\Delta(y-z)\delta(z-x)\delta(y'-x)\Delta(y'-z')J(z') = \\
 &= \frac{g}{48} \int d^4x d^4z' \Delta(x-x)\Delta(x-z')J(z') .
 \end{aligned}$$

This will correspond to:

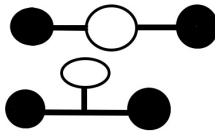


where the source  $J$  is located at the dot, and the vertex joining the line to the loop is at  $x$ .

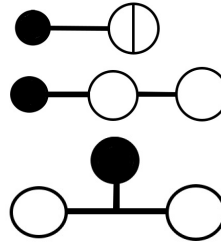
There are multiple possible diagrams for  $V = 1, P = 3$ ,



And for  $V = 2, P = 4$ ,



And for  $V = 3, P = 5$



and so on.

Through a series of combinatoric and physical arguments, it can be shown that only connected diagrams will contribute, and the  $1/P!$  and  $1/V!$  terms will always cancel exactly.

So, to calculate the amplitude for a particular interaction to happen (say  $N$   $\phi$ 's in and  $M$   $\phi$ 's out), draw every connected diagram that is topologically distinct, and has the correct number of *in* and *out* particles. Then, through a set of rules which you will learn formally in a QFT course, you can reconstruct the integrals which we started with in (9.44).

When you take a course on QFT, you will spend a tremendous amount of time learning how to evaluate these integrals for low order (they cannot be evaluated past about second order in most cases). While this is extremely important, it is not vital for the agenda of these notes, and we therefore do not discuss how they are evaluated.

The idea is that each diagram represents one of the possible paths the particle can take, along with the possible interactions it can be a part of. Because this is a quantum mechanical theory, we know it is actually in a superposition of all possible paths and interactions. We don't make a measurement or observation until the particles leave the area in which they collide, so we have no idea about what is going on inside the accelerator. We know that if *this* goes in and *this* comes out, we can draw a particular set of diagrams which have the correct input and output, and the nature of the interaction terms (which determines what types of vertices you can have) tells us what types of interactions we can have inside the accelerator. Evaluating the integrals then tells us how much that particular event/diagram contributes towards the total probability amplitude. So, if you want to know how likely a certain incoming/outgoing set of particles is, write down all the diagrams, evaluate the corresponding integrals, and add them up.



As we pointed out above, the classical behavior (which is more probable) is closer to the first order approximation of the quantum behavior. Therefore, even though in general we can't evaluate the integrals past about second order, the first few orders tell us to a reasonable (in fact, exceptional in most cases) degree of accuracy what the amplitude is. If we want more accuracy, we can seek to evaluate higher orders, but usually lower orders suffice for experiments at energy levels we can currently attain.

One of the difficulties encountered with evaluating these integrals is that you almost always find that they yield infinite amplitudes. Since an amplitude (which is a probability) should be between 0 and 1, this is obviously unacceptable. The process of finding the infinite parts and separating them from the finite parts of the amplitude is a very well defined mathematical construct called *Renormalization*. The basic idea is that any infinite term consists of a pure infinity and a finite part. For example, the infinite sum:

$$\sum_{n=1}^{\infty} n = \lim_{x \rightarrow 0} \frac{1}{x^2} - \frac{1}{12}. \quad (9.48)$$

There is a part which is a pure infinity (the first term on the right hand side), and a term which is finite. While this may seem strange and extremely unfamiliar (and a bit like hand waving), it is actually a very rigorous and very well understood mathematical idea.

Much of what particle physicists attempt to do is find theories (and types of theories) that can be renormalized and theories that cannot. For example, the action which leads to General Relativity leads to a quantum theory which cannot be renormalized. Renormalization is a fascinating and deep topic, and will be covered in great depth in any standard QFT text or course.

## 9.5 Interacting Fermion Fields

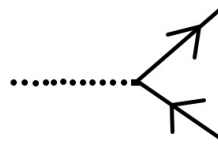
The analysis we performed above for scalar fields  $\phi$  above is almost identical for fermions, and we therefore won't repeat it. The main difference is that the interaction terms will have a field  $\psi$  interacting with  $\bar{\psi}$ , and so the vertices will be slightly different. We won't bother with those details.

Finally, we can have a Lagrangian with both scalars and fermions. Then, naturally, you could have interaction terms where the scalars interact with

fermions. While there are countless interaction terms of this type, the one that will be the most interesting to us is the *Yukawa* term,

$$\mathcal{L}_{\text{Yuk}} = g\phi\bar{\psi}\psi . \quad (9.49)$$

If we represent  $\phi$  by a dotted line,  $\psi$  by a line with an arrow in the forward time direction, and  $\bar{\psi}$  with an arrow going backwards in time, this interaction term will show up in a Feynman diagram as



Once each diagram is drawn, there are well defined rules to write down an integral corresponding to each diagram.

**Exercise 9.1:** Using the path integral representations find the transition amplitude for a free particle moving in one dimension.

**Exercise 9.2:** Find the time spreading of the initially ( $t = 0$ ) Gaussian Schrödinger wave packet

$$\Psi(q, 0) = \left( \frac{1}{2\pi\sigma^2} \right)^{1/4} e^{-\frac{(q-a)^2}{4\sigma^2}} ,$$

where  $a$  is some parameter and  $\sigma$  is a scalar field.

**Exercise 9.3:** Using the path integral find the transition amplitude for a harmonic oscillator.

**Exercise 9.4:** Find the path integral representation of the partition function for canonical ensemble.

## Part IV

# Lecture – Regularizations and Renormalizations



The relativistic field theory is full of infinities which need to be taken care of before the theoretical predictions can be compared with experimental measurements.

The problem of infinities first arose in the classical electrodynamics of point particles. The mass of a charged particle should include the mass-energy in its electrostatic field (electromagnetic mass). Assume that the particle is a charged spherical shell of radius  $r_e$ . The mass-energy in the field is

$$m_{\text{em}} = \int \frac{1}{2} E^2 dV = \int_{r_e}^{\infty} \frac{1}{2} \left( \frac{q}{4\pi r^2} \right)^2 4\pi r^2 dr = \frac{q^2}{8\pi r_e},$$

which becomes infinite as  $r_e \rightarrow 0$ . This implies that the point particle would have infinite inertia, making it unable to be accelerated. Incidentally, the value of  $r_e$  that makes  $m_{\text{em}}$  equal to the electron mass is called the classical electron radius,

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2} = \alpha \frac{\hbar}{m_e c} \approx 2.8 \times 10^{-15} \text{ m},$$

where  $\alpha \approx 1/137$  is the fine-structure constant, and  $\hbar/m_e c$  is the Compton wavelength of the electron.

The total effective mass of a spherical charged particle includes the actual bare mass of the spherical shell (in addition to the mass associated with its electric field). If the shell's bare mass is allowed to be negative, it might be possible to take a consistent point limit. This was called *renormalization*, and Lorentz and Abraham attempted to develop a classical theory of the electron this way.

When calculating the electromagnetic interactions of charged particles, it is tempting to ignore the back-reaction of a particle's own field on itself. But this back-reaction is necessary to explain the friction on charged particles when they emit radiation. If the electron is assumed to be a point, the value of the back-reaction diverges. The Abraham-Lorentz theory had a noncausal "pre-acceleration". Sometimes an electron would start moving before the force is applied. This early work was the inspiration for later attempts at *regularization* and *renormalization* in QFT, which are conceptually distinct concepts.

*Regularization* is the process by which one renders divergent quantities finite by introducing a parameter  $\Lambda$  such that the original divergent theory (the

theory which is ill-defined before regularization) corresponds to a certain value of that parameter. Once you regularize your theory, you can calculate any quantity you want in terms of the "bare" quantities appearing in the original lagrangian (such as masses  $m$ , couplings  $\lambda$ , etc.) along with the newly introduced regularization parameter  $\Lambda$ . The bare quantities are not what is measured in experiments. What is measured in experiments are corresponding physical quantities (the physical masses  $m_p$ , couplings  $\lambda_p$ , etc.).

*Renormalization* is the process by which you take the regularized theory, a theory written in terms of bare quantities and the regularization parameter  $(\Lambda, m, \lambda, \dots)$ , and you apply certain conditions (renormalization conditions) which cause physical quantities you want to compute, such as scattering amplitudes, to depend only on physical quantities  $(m_p, \lambda_p, \dots)$ , and in performing this procedure on a renormalizable QFT, the dependence on the cutoff disappears. So, in a sense, renormalization can be thought of as more of a procedure for writing your theory in terms of physical quantities than as a procedure for "removing infinities". The removing infinite part is already accomplished through regularization.

Eventhough the concept of renormalization is quite simple, the actual procedure for carrying out the operation is quite complicated and intimidating. We need to use some regularization procedure to make these divergent quantities finite before we can do mathematically meaningful manipulations. Note that not every relativistic field theory will have this property that all divergences can be absorbed into redefinition of few physical parameters. Those which have this property are called *renormalizable* theories and those which don't are called *unrenormalizable* theories. This has become an important criteria for choosing a right theory because we do not really know how to handle the unrenormalizable theory.

Beware that what I have described here is not the whole conceptual story of regularization and renormalization. More complete picture of how renormalization is thought about nowadays will be considered in this lecture.

## Chapter 10

# Regularization

In physics regularization is a method of modifying observables which have singularities in order to make them finite by the introduction of a suitable parameter called regulator. The regulator, also known as a "cutoff", models our lack of knowledge about physics at unobserved scales (e.g. scales of small size or large energy levels). It compensates for (and requires) the possibility that "new physics" may be discovered at those scales which the present theory is unable to model, while enabling the current theory to give accurate predictions as an "effective theory" within its intended scale of use.

Regularization was for many decades controversial even amongst its inventors, as it combines physical and epistemological claims into the same equations. However, it is now well understood and has proven to yield useful, accurate predictions.

Regularization procedures deal with infinite, divergent, and nonsensical expressions by introducing an auxiliary concept of a regulator (for example, the minimal distance in space, which is useful in case the divergences arise from short-distance physical effects). The correct physical result is obtained in the limit in which the regulator goes away, but the virtue of the regulator is that for its finite value, the result is finite.

Regularization is the first step towards obtaining a completely finite and meaningful result; in QFT it must be usually followed by a related, but independent technique called *Renormalization*. Renormalization is based on the requirement that some physical quantities are equal to the observed

values. Such a constraint allows one to calculate a finite value for many other quantities that looked divergent.

The existence of a limit and the independence of the final result from the regulator are nontrivial facts. The underlying reason for them lies in universality as shown by Wilson and Kadanoff considering the second order phase transitions. Sometimes, taking the limit is not possible. This is the case when we have a Landau pole and for nonrenormalizable couplings, like the Fermi interaction. However, working with scales of the order of regulators still give pretty accurate approximations. The physical reason why we can't take the limit is the existence of new physics.

It is not always possible to define a regularization such that the limit is independent of the regularization. In this case, one says that the theory contains an anomaly.

## 10.1 Realistic Regularization

Perturbative predictions by QFT are computed using the Feynman rules and a regularization method to circumvent UV divergences. The method results in regularized  $n$ -point Green's functions (propagators), and a suitable limiting procedure (a renormalization scheme) then leads to perturbative  $S$ -matrix elements. These are independent of the particular regularization method used, and enable one to model perturbatively the measurable physical processes (cross sections, probability amplitudes, decay widths and lifetimes of excited states). However, so far no known regularized  $n$ -point Green's functions can be regarded as being based on a physically realistic theory of quantum-scattering since the derivation of each disregards some of the basic tenets of conventional physics (e.g. by not being Lorentz-invariant, by introducing either unphysical particles with a negative metric or wrong statistics, or discrete space-time, or lowering the dimensionality of space-time, or some combination thereof, etc.). So the available regularization methods are understood as formalistic technical devices, devoid of any direct physical meaning.

As it seems that the vertices of non-regularized Feynman series adequately describe interactions in quantum scattering, it is taken that their UV divergences are due to the asymptotic, high-energy behavior of the Feynman propagators. So it is a prudent, conservative approach to retain the vertices



in Feynman series, and modify only the Feynman propagators to create a regularized Feynman series.

### 10.1.1 Historical Remarks and Opinions

In 1949 Pauli conjectured there is a realistic regularization, which is implied by a theory that respects all the established principles of contemporary physics. So its propagators:

- Do not need to be regularized;
- Can be regarded as such a regularization of the propagators used in QFTs that might reflect the underlying physics.

The additional parameters of such a theory do not need to be removed (i.e. no renormalization is needed) and may provide some new information about the physics of quantum scattering, though they may turn out experimentally to be negligible. By contrast, any present regularization method introduces formal coefficients that must eventually be disposed of by renormalization.

Dirac was persistently, extremely critical about procedures of renormalization. So in 1963 :"... in the renormalization theory we have a theory that has defied all the attempts of the mathematician to make it sound. I am inclined to suspect that the renormalization theory is something that will not survive in the future, ...". So he was expecting a realistic regularization.

Still is relevant Salam's remark (1972) about the skepticism on a realistic regularization: "Field-theoretic infinities first encountered in Lorentz's computation of electron have persisted in classical electrodynamics for seventy and in QED for some thirty-five years. These long years of frustration have left in the subject a curious affection for the infinities and a passionate belief that they are an inevitable part of nature; so much so that even the suggestion of a hope that they may after all be circumvented - and finite values for the renormalization constants computed - is considered irrational".

However, in 't Hooft's opinion: "History tells us that if we hit upon some obstacle, even if it looks like a pure formality or just a technical complication, it should be carefully scrutinized. Nature might be telling us something, and we should find out what it is".

By Dirac: "One can distinguish between two main procedures for a theoretical physicist. One of them is to work from the experimental basis ... . The other procedure is to work from the mathematical basis. One examines and criticizes the existing theory. One tries to pin-point the faults in it and then tries to remove them. The difficulty here is to remove the faults without destroying the very great successes of the existing theory".

The difficulty with a realistic regularization is that nothing could be destroyed by its bottom-up approach and it has not experimental basis.

Considering distinct theoretical problems, Dirac in 1963 suggested: "I believe separate ideas will be needed to solve these distinct problems and that they will be solved one at a time through successive stages in the future evolution of physics. At this point I find myself in disagreement with most physicists. They are inclined to think one master idea will be discovered that will solve all these problems together. I think it is asking too much to hope that anyone will be able to solve all these problems together. One should separate them one from another as much as possible and try to tackle them separately. And I believe the future development of physics will consist of solving them one at a time, and that after any one of them has been solved there will still be a great mystery about how to attack further ones".

According to Dirac: "Quantum electrodynamics is the domain of physics that we know most about, and presumably it will have to be put in order before we can hope to make any fundamental progress with other field theories, although these will continue to develop on the experimental basis".

Dirac's two preceding remarks suggest that we should start searching for a realistic regularization in the case of QED in the 4-dimensional Minkowski spacetime, starting with the original QED Lagrangian density.

The path-integral formulation provides the most direct way from the Lagrangian density to the corresponding Feynman series. The free-field part of the Lagrangian density determines the Feynman propagators, whereas the rest determines the vertices. As the QED vertices are considered to adequately describe interactions, it makes sense to modify only the free-field part of the Lagrangian density so as to obtain such regularized Feynman series that the Lehmann-Symanzik-Zimmermann reduction formula provides a perturbative  $S$ -matrix that: (a) Is Lorentz invariant and unitary; (b) Involves only the QED particles; (c) Depends solely on QED parameters and those introduced by the modification of the Feynman propagators; (d) Exhibits the

same symmetries as the QED perturbative  $S$ -matrix. Let us refer to such a regularization as the *minimal realistic regularization*, and start searching for the corresponding, modified free-field parts of the QED Lagrangian density.

According to Bjorken and Drell, it would make physical sense to sidestep UV divergences by using more detailed description than can be provided by differential field equations. Feynman noted about the use of differential equations: "... for neutron diffusion it is only an approximation that is good when the distance over which we are looking is large compared with the mean free path. If we looked more closely, we would see individual neutrons running around". Then he wondered, "Could it be that the real world consists of little  $X$ -ons which can be seen only at very tiny distances? And that in our measurements we are always observing on such a large scale that we can't see these little  $X$ -ons, and that is why we get the differential equations? ... Are they also correct only as a smoothed-out imitation of a really much more complicated microscopic world?"

Already in 1938, Heisenberg proposed that a QFT can provide only a large-scale description of quantum dynamics, valid for distances larger than some fundamental length, expected also by Bjorken and Drell in 1965. Feynman's preceding remark provides a possible physical reason for its existence; either that or it is just another way of saying the same thing (there is a fundamental unit of distance) but having no new information.

The need for regularizations in any quantum theory of gravity is one of the major motivation for physics beyond SM. Infinities of the non-gravitational forces in QFT can be controlled via renormalization only, but additional regularization (and hence new physics) is required uniquely for gravity. The regularizers model break down of QFT at small scales and thus show clearly the need for some other theory to come into play at these scales. Zee (2003) considers this to be a benefit of the regularization framework – theories can work well in their intended domains but also contain information about their own limitations and point clearly to where new physics is needed.

## 10.2 Dimensional Regularization

*Dimensional regularization* is a method introduced by Giambiagi and Bollini, and independently and more comprehensively by 't Hooft and Veltman, for regularizing integrals in the evaluation of Feynman diagrams, by assigning

values to them that are meromorphic functions of a complex parameter  $d$ , the analytic continuation of the number of spacetime dimensions.

Dimensional regularization writes a Feynman integral as an integral depending on the spacetime dimension  $d$  and the squared distances  $(x_i - x_j)^2$  of the spacetime points  $x_i$  appearing in it.

In Euclidean space, the integral often converges for  $Re(d)$  sufficiently large, and can be analytically continued from this region to a meromorphic function defined for all complex  $d$ . In general, there will be a pole at the physical value (usually 4) of  $d$ , which needs to be canceled by renormalization to obtain physical quantities. Etingof (1999) showed that dimensional regularization is mathematically well defined, at least in the case of massive Euclidean fields, by using the Bernstein-Sato polynomial to carry out the analytic continuation.

Although the method is most well understood when poles are subtracted and  $d$  is once again replaced by 4, it has also led to some successes when  $d$  is taken to approach another integer value where the theory appears to be strongly coupled, as in the case of the Wilson-Fisher fixed point. A further leap is to take the interpolation through fractional dimensions seriously. This has led some authors to suggest that dimensional regularization can be used to study the physics of crystals that macroscopically appear to be fractals.

If one wishes to evaluate a loop integral which is logarithmically divergent in four dimensions, like

$$\int \frac{d^d p}{(2\pi)^d} \frac{1}{(p^2 + m^2)^2}, \quad (10.1)$$

one first rewrites the integral in some way so that the number of variables integrated over does not depend on  $d$ , and then we formally vary the parameter  $d$ , to include non-integral values like  $d = 4 - \epsilon$ . This gives

$$\begin{aligned} \int_0^\infty \frac{dp}{(2\pi)^{4-\epsilon}} \frac{2\pi^{(4-\epsilon)/2}}{\Gamma\left(\frac{4-\epsilon}{2}\right)} \frac{p^{3-\epsilon}}{(p^2 + m^2)^2} &= \frac{2^{\epsilon-4}\pi^{\frac{\epsilon}{2}-1}}{\sin\left(\frac{\pi\epsilon}{2}\right)\Gamma\left(1 - \frac{\epsilon}{2}\right)} m^{-\epsilon} = \\ &= \frac{1}{8\pi^2\epsilon} - \frac{1}{16\pi^2} \left( \ln \frac{m^2}{4\pi} + \gamma \right) + \mathcal{O}(\epsilon), \end{aligned} \quad (10.2)$$

where  $\gamma = 0.5772\dots$ .

### 10.3 Pauli-Villars Regularization

*Pauli-Villars regularization* is a procedure that isolates divergent terms from finite parts in loop calculations in QFT in order to renormalize the theory. Pauli and Villars published the method in 1949, based on earlier work by Feynman, Stueckelberg and Rivier.

In this treatment, a divergence arising from a loop integral (such as vacuum polarization or electron self-energy) is modulated by a spectrum of auxiliary particles added to the Lagrangian or propagator. When the masses of the fictitious particles are taken as an infinite limit (i.e. once the regulator is removed) one expects to recover the original theory.

This regulator is gauge invariant due to the auxiliary particles being minimally coupled to the photon field through the gauge covariant derivative. In QCD calculations it is not gauge covariant, so Pauli-Villars regularization cannot be used and serves as an alternative to the more favorable dimensional regularization in specific circumstances, such as in chiral phenomena, where a change of dimension alters the properties of the Dirac  $\gamma$ -matrices.

Using the path-integral formalism, 't Hooft and Veltman had invented the method of unitary regulators, which is a Lagrangian based Pauli-Villars method with a discrete spectrum of auxiliary masses.

Pauli-Villars regularization consists of introducing a fictitious mass term. For example, we would replace a photon propagator

$$\frac{1}{k^2 + i\epsilon} , \tag{10.3}$$

by

$$\frac{1}{k^2 + i\epsilon} - \frac{1}{k^2 - M^2 + i\epsilon} , \tag{10.4}$$

where  $M$  can be thought of as the mass of a fictitious heavy photon, whose contribution is subtracted from that of an ordinary photon.

### 10.4 Lattice Regularization

Lattice field theory is the study of lattice models of QFT, that is, of field theory on a spacetime that has been discretized onto a lattice. Although most

lattice field theories are not exactly solvable, they are of tremendous appeal because they can be studied by simulation on a computer. One hopes that, by performing simulations on larger and larger lattices, while making the lattice spacing smaller and smaller, one will be able to recover the behavior of the continuum theory.

Just as in all lattice models, numerical simulation gives access to field configurations that are not accessible to perturbation theory, such as solitons. Likewise, non-trivial vacuum states can be discovered and probed.

The method is particularly appealing for the quantization of a gauge theory. Most quantization methods keep Poincaré invariance manifest but sacrifice manifest gauge symmetry by requiring gauge fixing. Only after renormalization can gauge invariance be recovered. Lattice field theory differs from these in that it keeps manifest gauge invariance, but sacrifices manifest Poincaré invariance – recovering it only after renormalization.

## 10.5 Zeta Function Regularization

*Zeta function regularization* is a type of summability method that assigns finite values to divergent sums or products, and in particular can be used to define determinants and traces of some self-adjoint operators. The technique is now commonly applied to problems in physics, but has its origins in attempts to give precise meanings to ill-conditioned sums appearing in number theory.

There are several different summation methods called zeta function regularization for defining the sum of a possibly divergent series  $a_1 + a_2 + \dots$ .

- One method is to define its zeta regularized sum

$$\zeta_A(s) = \frac{1}{a_1^s} + \frac{1}{a_2^s} + \dots \quad (10.5)$$

where the zeta function is defined for  $Re(s)$  and by analytic continuation elsewhere. In the case when  $a_n = n$ , the zeta function is the ordinary Riemann zeta function, and this method was used by Euler to "sum" the series  $1 + 2 + 3 + 4 + \dots$  to  $\zeta(-1) = -1/12$ .

Other values of  $s$  can also be used to assign for the divergent sums

$$1+1+1+1+\dots \rightarrow \zeta(0) = -1/2, \quad 1+4+9+\dots \rightarrow \zeta(-2) = 0, \quad (10.6)$$

and in general

$$\sum_{n=1}^{\infty} n^s = 1^s + 2^s + 3^s + \dots \rightarrow \zeta(-s) = -\frac{B_{s+1}}{s+1}, \quad (10.7)$$

where  $B_k$  is a Bernoulli number.

Hawking showed (1977) that in flat space, in which the eigenvalues of Laplacians are known, the zeta function corresponding to the partition function can be computed explicitly. Consider a scalar field  $\phi$  contained in a large box of volume  $V$  in flat spacetime at the temperature  $T = \beta^{-1}$ . The partition function is defined by a path integral over all fields  $\phi$  on the Euclidean space obtained by putting  $\tau = it$ , which are zero on the walls of the box and which are periodic in  $\tau$  with period  $\beta$ . In this situation from the partition function he computes energy, entropy and pressure of the radiation of the field  $\phi$ . In case of flat spaces the eigenvalues appearing in the physical quantities are generally known, while in case of curved space they are not known: in this case asymptotic methods are needed.

- Another method defines the possibly divergent infinite product  $a_1, a_2, \dots$  to be  $e^{-\zeta' A(0)}$ . Ray and Singer (1971) used this to define the determinant of a positive self-adjoint operator  $A$  (the Laplacian of a Riemannian manifold in their application) with eigenvalues  $a_1, a_2, \dots$ , and in this case the zeta function is formally the trace of  $A^{-s}$ . Minakshisundaram and Pleijel (1949) showed that if  $A$  is the Laplacian of a compact Riemannian manifold then the Minakshisundaram-Pleijel zeta function converges and has an analytic continuation as a meromorphic function to all complex numbers, and Seeley (1967) extended this to elliptic pseudo-differential operators  $A$  on compact Riemannian manifolds. So for such operators one can define the determinant using zeta function regularization.

Hawking (1977) suggested using this idea to evaluate path integrals in curved spacetimes. He studied zeta function regularization in order to calculate the partition functions for thermal graviton and matter's quanta in curved background such as on the horizon of black holes and on de Sitter background using the relation by the inverse Mellin transformation to the trace of the kernel of heat equations.

The first example in which zeta function regularization is available appears in the Casimir effect, which is in a flat space with the bulk contributions of the quantum field in three space dimensions. In this case we must calculate the value of Riemann zeta function at  $-3$ , which diverges explicitly. However, it can be analytically continued to  $s = -3$  where hopefully there is no pole, thus giving a finite value to the expression. A detailed example of this regularization at work is given in the article on the detail example of the Casimir effect, where the resulting sum is very explicitly the Riemann zeta-function (and where the seemingly legerdemain analytic continuation removes an additive infinity, leaving a physically significant finite number).

An example of zeta-function regularization is the calculation of the vacuum expectation value of the energy of a particle field in quantum field theory. More generally, the zeta-function approach can be used to regularize the whole energy-momentum tensor in curved spacetime.

The unregulated value of the energy is given by a summation over the zero-point energy of all of the excitation modes of the vacuum:

$$\langle 0|T_{00}|0\rangle = \sum_n \frac{\hbar|\omega_n|}{2} . \quad (10.8)$$

Here, the sum (which may be an integral) is understood to extend over all energy modes  $\pm\omega_n$ ; the absolute value reminding us that the energy is taken to be positive. This sum, as written, is usually infinite ( $\omega_n$  is typically linear in  $n$ ). The sum may be regularized by writing it as

$$\langle 0|T_{00}(s)|0\rangle = \sum_n \frac{\hbar|\omega_n|}{2} |\omega_n|^{-s} , \quad (10.9)$$

where  $s$  is some parameter, taken to be a complex number. For large, real  $s$  greater than 4 (for three-dimensional space), the sum is manifestly finite, and thus may often be evaluated theoretically.

The zeta-regularization is useful as it can often be used in a way such that the various symmetries of the physical system are preserved. Zeta-function regularization is used in conformal field theory, renormalization and in fixing the critical spacetime dimension of string theory.

We can ask if are there any relation to the dimensional regularization originated by the Feynman diagram. But now we may say they are equivalent



each other. However the main advantage of the zeta regularization is that it can be used whenever the dimensional regularization fails, for example if there are matrices or tensors inside the calculations.

### 10.5.1 Relation to Dirichlet Series

Zeta-function regularization gives an analytic structure to any sums over an arithmetic function  $f(n)$ . Such sums are known as Dirichlet series. The regularized form,

$$\tilde{f}(s) = \sum_{n=1}^{\infty} f(n)n^{-s} , \quad (10.10)$$

converts divergences of the sum into simple poles on the complex  $s$ -plane. In numerical calculations, the zeta-function regularization is inappropriate, as it is extremely slow to converge. For numerical purposes, a more rapidly converging sum is the exponential regularization, given by

$$F(t) = \sum_{n=1}^{\infty} f(n)e^{-tn} . \quad (10.11)$$

This is sometimes called the  $Z$ -transform of  $f$ , where  $Z = e^{-t}$ . The analytic structure of the exponential and zeta-regularizations are related. By expanding the exponential sum as a Laurent series

$$F(t) = \frac{a_N}{t^N} + \frac{a_{N-1}}{t^{N-1}} + \dots \quad (10.12)$$

one finds that the zeta-series has the structure

$$\tilde{f}(s) = \frac{a_N}{s - N} + \dots . \quad (10.13)$$

The structure of the exponential and zeta-regularizers are related by means of the Mellin transform. The one may be converted to the other by making use of the integral representation of the Gamma function:

$$\Gamma(s + 1) = \int_0^{\infty} x^s e^{-x} dx , \quad (10.14)$$

which lead to the identity

$$\Gamma(s + 1)\tilde{f}(s + 1) = \int_0^{\infty} t^s F(t) dt , \quad (10.15)$$

relating the exponential and zeta-regularizers, and converting poles in the  $s$ -plane to divergent terms in the Laurent series.

### 10.5.2 Heat Kernel Regularization

Much of the early work establishing the convergence and equivalence of series regularized with the heat kernel and zeta function regularization methods was done by Hardy and Littlewood in 1916 and is based on the application of the Cahen-Mellin integral. The effort was made in order to obtain values for various ill-defined, conditionally convergent sums appearing in number theory.

The sum

$$f(s) = \sum_n a_n e^{-s|\omega_n|} \quad (10.16)$$

is sometimes called a heat kernel or a heat-kernel regularized sum; this name stems from the idea that the  $\omega_n$  can sometimes be understood as eigenvalues of the heat kernel. In mathematics, such a sum is known as a generalized Dirichlet series; its use for averaging is known as an Abelian mean. It is closely related to the Laplace-Stieltjes transform, in that

$$f(s) = \int_0^\infty e^{-st} d\alpha(t) , \quad (10.17)$$

where  $\alpha(t)$  is a step function, with steps of  $a_n$  at  $t = |\omega_n|$ . A number of theorems for the convergence of such a series exist. For example, by the Hardy-Littlewood Tauberian theorem, if

$$L = \limsup_{n \rightarrow \infty} \frac{\log |\sum_{k=1}^n a_k|}{|\omega_n|} \quad (10.18)$$

then the series for  $f(s)$  converges in the half-plane  $\Re(s) > L$  and is uniformly convergent on every compact subset of the half-plane  $\Re(s) > L$ . In almost all applications to physics, one  $L = 0$ .

## 10.6 Causal Perturbation Theory

*Causal perturbation theory*, which goes back to a work by Epstein and Glaser (1973), is a mathematically rigorous approach to renormalization theory, which makes it possible to put the theoretical setup of perturbative QFT on a sound mathematical basis.

When developing QED in the 1940s, Tomonaga, Schwinger, Feynman and Dyson discovered that, in perturbative calculations, problems with divergent integrals abounded. The divergences appeared in calculations involving Feynman diagrams with closed loops of virtual particles. It is an important observation that time-ordered products of distributions arise in a natural way and may lead to UV divergences in the corresponding calculations. From the mathematical point of view, the problem of divergences is rooted in the fact that the theory of distributions is a purely linear theory, in the sense that the product of two distributions cannot consistently be defined (in general), as was proved by Schwartz in the 1950s.

Epstein and Glaser solved this problem for a special class of distributions that fulfill a causality condition, which itself is a basic requirement in axiomatic QFT, they studied only theories involving scalar particles. Since then, the causal approach has been applied also to a wide range of gauge theories, which represent the most important QFTs in modern physics.

## 10.7 Hadamard Regularization

*Hadamard regularization* (also called Hadamard finite part) is a method of regularizing divergent integrals by dropping some divergent terms and keeping the finite part, introduced by Hadamard (1923, 1932). Riesz (1938, 1949) showed that this can be interpreted as taking the meromorphic continuation of a convergent integral.

If the Cauchy principal value integral

$$\mathcal{C} \int_a^b \frac{f(t)}{t-x} dt \quad (a < x < b) \quad (10.19)$$

exists, then it may be differentiated with respect to  $x$  to obtain the Hadamard finite part integral as follows:

$$\frac{d}{dx} \left( \mathcal{C} \int_a^b \frac{f(t)}{t-x} dt \right) = \mathcal{H} \int_a^b \frac{f(t)}{(t-x)^2} dt . \quad (a < x < b) \quad (10.20)$$

Note that the symbols  $\mathcal{C}$  and  $\mathcal{H}$  are used here to denote Cauchy principal value and Hadamard finite-part integrals respectively.

The Hadamard finite part integral above (for  $a < x < b$ ) may also be given

by the following equivalent definitions:

$$\begin{aligned} \mathcal{H} \int_a^b \frac{f(t)}{(t-x)^2} dt &= \lim_{\varepsilon \rightarrow 0^+} \left\{ \int_a^{x-\varepsilon} \frac{f(t)}{(t-x)^2} dt + \int_{x+\varepsilon}^b \frac{f(t)}{(t-x)^2} dt - \frac{2f(x)}{\varepsilon} \right\} = \\ &= \lim_{\varepsilon \rightarrow 0^+} \left\{ \int_a^b \frac{(t-x)^2 f(t)}{(t-x)^2 + \varepsilon^2} dt - \frac{\pi f(x)}{2\varepsilon} - \frac{f(x)}{2} \left( \frac{1}{b-x} - \frac{1}{a-x} \right) \right\}. \end{aligned} \quad (10.21)$$

The definitions above may be derived by assuming that the function  $f(t)$  is differentiable infinitely many times at  $t = x$  ( $a < x < b$ ), that is, by assuming that  $f(t)$  can be represented by its Taylor series about  $t = x$ .

Integral equations containing Hadamard finite part integrals (with  $f(t)$  unknown) are termed hypersingular integral equations. Hypersingular integral equations arise in the formulation of many problems in mechanics, such as in fracture analysis.

**Exercise 10.1:** Calculate the Feynman integral in  $n$  dimensions:

$$\int \frac{d^n k}{(2\pi)^n} \frac{1}{(k^2 + 2p \cdot k + m^2 + i\epsilon)^2},$$

using the dimensional regularization.

**Exercise 10.2:** Use the dimensional regularization to compute the one-loop vacuum polarization in QED.

**Exercise 10.3:** In scalar electrodynamics two diagrams give contribution to the polarization of vacuum. Using dimensional regularization derive the expression for the divergent part of the vacuum polarization.

## Chapter 11

# Renormalization

*The Theory of Renormalization* is a prescription which consistently isolates and removes all infinities from the physically measurable quantities. Note that the need for renormalization is quite general and is not unique to the relativistic field theory. For example, consider an electron moving inside a solid. If the interaction between electron and the lattice of the solid is weak enough, we can use an effective mass  $m^*$  to describe its response to an externally applied force and this effective mass is certainly different from the mass  $m$  measured outside the solid. Thus the electron mass is changed (renormalized) from  $m$  to  $m^*$  by the interaction of the electron with the lattice in the solid. In this simple case, both  $m$  and  $m^*$  are measurable and hence finite. For the relativistic field theory, the situation is the same except for two important differences:

1. The renormalization due to the interaction is generally infinite (corresponding to the divergent loop diagrams). These infinities, coming from the contribution of high momentum modes are present even for the cases where the interactions are weak;
2. There is no way to switch off the interaction between particles and the quantities in the absence of interaction – bare quantities are not measurable. Roughly speaking, the program of removing the infinities from physically measurable quantities involves shuffling all the divergences into bare quantities. In other words, we can redefine the unmeasurable quantities to absorb the divergences so that the physically measurable quantities are finite. The renormalized mass which is now finite can

only be determined from experimental measurement and cannot be predicted from the theory alone.

There are two different methods to carry out the renormalization program: *Conventional Renormalization*, which is more intuitive but mathematically complicated, and *BPH Renormalization*, which is simple to describe but not so transparent. These two methods are in fact complementary to each other and it is very useful to know both.

## 11.1 Conventional Renormalization

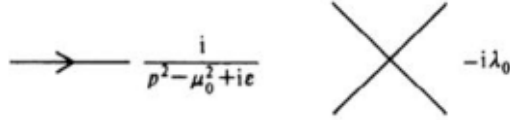
We will illustrate this scheme in the simple  $\lambda\varphi^4$  theory where the Lagrangian can be written as:

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 , \quad (11.1)$$

with

$$\mathcal{L}_0 = \frac{1}{2} [(\partial_\mu \varphi_0)^2 - \mu_0^2 \varphi_0^2] , \quad \mathcal{L}_1 = -\frac{\lambda_0}{4!} \varphi_0^4 . \quad (11.2)$$

Here  $\mu_0$ ,  $\lambda_0$ ,  $\varphi_0$  are bare mass, bare coupling constant and bare field, respectively. The propagator and vertex of this theory are given below,



Here  $p$  is the momentum carried by the line and  $\mu_0^2$  is the bare mass in  $\mathcal{L}_0$ .

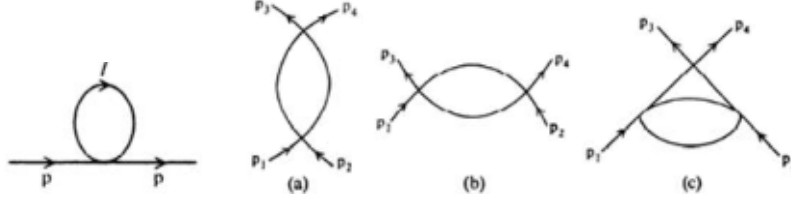
The two point function (propagator) defined by

$$i\Delta(p) = \int d^4x e^{-ip \cdot x} \langle 0 | T(\varphi_0(x) \varphi_0(0)) | 0 \rangle , \quad (11.3)$$

can be written in terms of one-particle-irreducible (1PI), using those graphs which cannot be made disconnected by cutting any one internal line, i.e. can be written as a geometric series

$$\begin{aligned} i\Delta(p) &= \frac{i}{p^2 - \mu_0^2 + i\varepsilon} + \frac{i}{p^2 - \mu_0^2 + i\varepsilon} (-i\Sigma(p^2)) \frac{i}{p^2 - \mu_0^2 + i\varepsilon} + \dots = \\ &= \frac{i}{p^2 - \mu_0^2 - \Sigma(p^2) + i\varepsilon} . \end{aligned} \quad (11.4)$$

Here  $\Sigma(p^2)$  is the 1PI self-energy graph. In one-loop, the 1PI 2-point and 4-point divergent graphs are:



For the self-energy the contribution is,

$$-i\Sigma(p^2) = -\frac{i\lambda_0}{2} \int \frac{d^4l}{(2\pi)^4} \frac{i}{l^2 - \mu_0^2 + i\varepsilon}, \quad (11.5)$$

which diverges quadratically and for the 4-point functions we have

$$\begin{aligned} \Gamma_a &= \Gamma(s) = \frac{(-i\lambda_0)^2}{2} \int \frac{d^4l}{(2\pi)^4} \frac{i}{l^2 - \mu_0^2 + i\varepsilon} \frac{i}{(l-p)^2 - \mu_0^2 + i\varepsilon}, \\ \Gamma_b &= \Gamma(t), \quad \Gamma_c = \Gamma(u), \end{aligned} \quad (11.6)$$

where  $s$ ,  $t$  and  $u$  are the Mandelstam variables (see Sec. 4.2.1) and  $\Gamma(s)$  diverges logarithmically.

Important feature to note about these integrals is that when we differentiate them with respect to external momenta, the integral will become more convergent. For example, if we differentiate  $\Gamma(p^2)$  by  $p^2$ , one finds

$$\begin{aligned} \frac{\partial}{\partial p^2} \Gamma(p^2) &= \frac{1}{2p^2} p_\mu \frac{\partial}{\partial p_\mu} \Gamma(p^2) = \\ &= \frac{\lambda_0^2}{p^2} \int \frac{d^4l}{(2\pi)^4} \frac{(l-p) \cdot p}{l^2 - \mu_0^2 + i\varepsilon} \frac{1}{[(l-p)^2 - \mu_0^2 + i\varepsilon]^2}, \end{aligned} \quad (11.7)$$

which is finite. This means that the divergences will reside only in the first few terms in a Taylor expansion in the external momenta of the Feynman diagram. In our case, we can write

$$\Gamma(s) = \Gamma(0) + \bar{\Gamma}(s), \quad (11.8)$$

where  $\Gamma(0)$  is logarithmic divergent and  $\bar{\Gamma}(s)$ , which is the sum of all higher derivative terms, is finite. In other words, the finite part  $\bar{\Gamma}(s)$  corresponds to subtracting the divergent part  $\Gamma(0)$  from  $\Gamma(s)$  and is sometimes referred to as the *subtraction*.

### 11.1.1 Mass and Wavefunction Renormalization

The self-energy contribution in (11.5) is quadratically divergent. To isolate the divergences we use the Taylor expansion around some arbitrary value  $\mu^2$ ,

$$\Sigma(p^2) = \Sigma(\mu^2) + (p^2 - \mu^2) \Sigma'(\mu^2) + \tilde{\Sigma}(p^2) , \quad (11.9)$$

where  $\Sigma(\mu^2)$  is quadratically divergent, and  $\Sigma'(\mu^2)$  is logarithmically divergent and  $\tilde{\Sigma}(p^2)$  is finite. The finite part  $\tilde{\Sigma}(p^2)$  will have the property,

$$\tilde{\Sigma}(\mu^2) = \tilde{\Sigma}'(\mu^2) = 0 . \quad (11.10)$$

Note that self-energy in 1-loop has the peculiar feature that it is independent of the external momentum  $p^2$  and the Taylor expansion has only one term,  $\Sigma(\mu^2)$ . However, the higher loop contribution depend on  $p^2$  and the Taylor expansion is non-trivial. The propagator in (11.4) is then,

$$i\Delta(p) = \frac{i}{p^2 - \mu_0^2 - \Sigma(\mu^2) - (p^2 - \mu^2) \Sigma'(\mu^2) - \tilde{\Sigma}(p^2) + i\varepsilon} . \quad (11.11)$$

The physical mass is defined as the position of the pole in the propagator. Since up to this point  $\mu^2$  is arbitrary, we can choose it to satisfy the relation,

$$\mu_0^2 + \Sigma(\mu^2) = \mu^2 . \quad (11.12)$$

Then

$$i\Delta(p) = \frac{i}{(p^2 - \mu^2) [1 - \Sigma'(\mu^2)] - \tilde{\Sigma}(p^2) + i\varepsilon} , \quad (11.13)$$

and using (11.10) we see that  $\Delta(p)$  has a pole at  $p^2 = \mu^2$ . Thus  $\mu^2$  is *the physical mass* and is related to the bare mass  $\mu_0^2$  in (11.12). This is *the mass renormalization*. Since  $\Sigma(\mu^2)$  is divergent, the bare mass  $\mu_0^2$  must also be divergent so that the combination  $\mu_0^2 + \Sigma(\mu^2)$  is finite and measurable. In other words, the bare mass  $\mu_0^2$  has to diverge in such a way that its divergence cancels the divergent loop correction to yield a finite result. It amounts to shuffling the infinities to unobservable quantities like bare mass  $\mu_0^2$ . This is the part in renormalization theory which is very difficult to comprehend at the first sight. Nevertheless it is logically consistent and the rules are very precise. Furthermore, the results after the renormalization have been successfully checked by experiments. This gives us confidence about the validity of renormalization theory.



To remove the divergent quantity  $\Sigma'(\mu^2)$  we note that in 1-loop both  $\Sigma'(\mu^2)$ ,  $\tilde{\Sigma}(p^2)$  are of order  $\lambda_0$ , for convenience, we can make the approximation,

$$\tilde{\Sigma}(p^2) \simeq [1 - \Sigma'(\mu^2)] \tilde{\Sigma}(p^2) + O(\lambda_0^2) , \quad (11.14)$$

and write the propagator as

$$i\Delta(p) = \frac{iZ_\varphi}{(p^2 - \mu^2) - \tilde{\Sigma}(p^2) + i\varepsilon} , \quad (11.15)$$

where

$$Z_\varphi = [1 - \Sigma'(\mu^2)]^{-1} \simeq 1 + \Sigma'(\mu^2) + O(\lambda_0^2) . \quad (11.16)$$

Now the divergence is shuffled into the multiplicative factor  $Z_\varphi$  which can be removed by defining a renormalized field  $\varphi$  as,

$$\varphi = Z_\varphi^{-1/2} \varphi_0 . \quad (11.17)$$

The propagator for the renormalized field is then

$$\begin{aligned} i\Delta_R(p) &= \int d^4x e^{-ip \cdot x} \langle 0 | T(\varphi(x) \varphi(0)) | 0 \rangle = \\ &= iZ_\varphi^{-1} \Delta(p) = \frac{i}{(p^2 - \mu^2) - \tilde{\Sigma}(p^2) + i\varepsilon} , \end{aligned} \quad (11.18)$$

and it is completely finite.  $Z_\varphi$  is usually called the *wavefunction renormalization constant*. Thus another divergence is shuffled into the bare field operator  $\varphi_0$  which is also not measurable.

The new renormalized field operator  $\varphi$  should also be applied to the renormalized higher point Green's functions,

$$\begin{aligned} G_R^{(n)}(x_1, x_2, \dots, x_n) &= \langle 0 | T(\varphi(x_1) \varphi(x_2) \dots \varphi(x_n)) | 0 \rangle = \\ &= Z_\varphi^{-n/2} \langle 0 | T(\varphi_0(x_1) \varphi_0(x_2) \dots \varphi_0(x_n)) | 0 \rangle = (11.19) \\ &= Z_\varphi^{-n/2} G_0^{(n)}(x_1, x_2, \dots, x_n) . \end{aligned}$$

Here  $G_0^{(n)}(x_1, x_2, \dots, x_n)$  is the unrenormalized  $n$ -point Green's function. Or in momentum space

$$G_R^{(n)}(p_1, p_2, \dots, p_n) = Z_\varphi^{-n/2} G_0^{(n)}(p_1, p_2, \dots, p_n) , \quad (11.20)$$

where

$$\begin{aligned} & (2\pi)^4 \delta^4(p_1 + \cdots p_n) G_R^{(n)}(p_1, \cdots p_n) = \\ & = \int \left( \prod_{i=1}^n dx_i^4 e^{-ip_i \cdot x_i} \right) G_R^{(n)}(x_1, \cdots x_n) . \end{aligned} \quad (11.21)$$

Similarly for  $G_0^{(n)}(p_1, p_2, \cdots p_n)$ . To go from the connected Green's functions to the 1PI (amputated) Green's functions, we need to eliminate the one-particle reducible diagrams, and also to remove the propagators  $i\Delta_R(p_i)$  for the external lines in 1PI Green's function  $G_R^{(n)}(p_1, \cdots p_n)$ . As a result the relation between 1PI Green's functions are of the form,

$$\Gamma_R^{(n)}(p_1, p_2, \cdots p_n) = Z_\varphi^{n/2} \Gamma_0^{(n)}(p_1, p_2, \cdots p_n) . \quad (11.22)$$

Note that the relations in (11.10) are direct consequence of the Taylor expansion around the point  $p^2 = \mu^2$  which is totally arbitrary. From the form of the renormalized propagator in (11.18), we see that (11.10) are equivalent to the relations

$$\Delta_R^{-1}(\mu^2) = 0 , \quad \left. \frac{d}{dp^2} \Delta_R^{-1}(p^2) \right|_{p^2=\mu^2} = 1 . \quad (11.23)$$

If we have chosen some other point, e.g.  $p^2 = 0$  for the Taylor expansion, the finite part  $\tilde{\Sigma}_1(p^2)$  will have the properties

$$\tilde{\Sigma}_1(0) = \tilde{\Sigma}'_1(0) = 0 . \quad (11.24)$$

Or in terms of renormalized propagator,

$$\Delta_R^{-1}(0) = -\mu^2, \quad \left. \frac{d}{dp^2} \Delta_R^{-1}(p^2) \right|_{p^2=0} = 1 . \quad (11.25)$$

Sometimes in the renormalization prescription we replace the statement *Taylor expansion around  $p^2 = \mu^2$ , or  $p^2 = 0$*  by relations expressed, in (11.10) and (11.24), called the *renormalization conditions*. One important feature to keep in mind is that in carrying out the renormalization program there is an arbitrariness in choosing the points for the Taylor expansion. Different renormalization schemes seem to give rise to different looking relations. However, if these renormalization schemes make any sense at all, the physical laws which are relations among physically measurable quantities should be the same regardless of which scheme is used. This is the basic idea behind the *renormalization group equations*.

### 11.1.2 Coupling Constant Renormalization

The basic coupling in  $\lambda\varphi^4$  theory is the 4-point function which in 1-loop has the form, before renormalization,

$$\Gamma_0^{(4)}(s, t, u) = -i\lambda_0 + \Gamma(s) + \Gamma(t) + \Gamma(u) , \quad (11.26)$$

where last three terms are logarithmic divergent. We will remove these divergences by the redefinition of the coupling constant. Note that the physical coupling constant is measured in terms of two-particle scattering amplitude which is essentially 1PI 4-point Green's function  $\Gamma_R^{(4)}(s, t, u)$  which is a function of the kinematical variables,  $s, t$  and  $u$ . For convenience, we can choose the symmetric point,

$$s_0 = t_0 = u_0 = \frac{4\mu^2}{3} \quad (11.27)$$

to define the coupling constant,

$$\Gamma_R^{(4)}(s_0, t_0, u_0) = -i\lambda , \quad (11.28)$$

where  $\lambda$  is the renormalized coupling constant. Since  $\Gamma(s)$  is only logarithmically divergent, we can isolate the divergence in one term in the Taylor expansion,

$$\Gamma(s) = \Gamma(s_0) + \tilde{\Gamma}(s) , \quad (11.29)$$

where  $\tilde{\Gamma}(s)$  is finite and  $\tilde{\Gamma}(s_0) = 0$ . Then

$$\Gamma_0^{(4)}(s, t, u) = -i\lambda_0 + 3\Gamma(s_0) + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) . \quad (11.30)$$

We can isolate the divergence by combining the first two term and define the vertex renormalization constant  $Z_\lambda$  by

$$-iZ_\lambda\lambda_0 = -i\lambda_0 + 3\Gamma(s_0) . \quad (11.31)$$

Then

$$\Gamma_R^{(4)}(s, t, u) = -iZ_\lambda\lambda_0 + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) . \quad (11.32)$$

The renormalized 4-point 1PI is then

$$\begin{aligned} \Gamma_R^{(4)}(s, t, u) &= Z_\varphi^2 \Gamma_0^{(4)}(s, t, u) = \\ &= -iZ_\lambda Z_\varphi^2 \lambda_0 + Z_\varphi^2 \left[ \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) \right] . \end{aligned} \quad (11.33)$$

We now define the renormalized coupling constant  $\lambda$  as

$$\lambda = Z_\lambda Z_\varphi^2 \lambda_0 , \quad (11.34)$$

and from (11.16) we see that

$$Z_\varphi = 1 + O(\lambda_0) . \quad (11.35)$$

Also  $\tilde{\Gamma}$  is of order of  $\lambda_0^2$ . The renormalized 4-point 1PI can be put into the form,

$$\Gamma_R^{(4)}(s, t, u) = \lambda + \left[ \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) \right] + O(\lambda_0^3) . \quad (11.36)$$

Assuming that the coupling constant  $\lambda$  is measured in the scattering and is finite, we see that this 4-point function is completely free of divergences. Equation (11.34) shows that the renormalization of coupling constant involves wavefunction renormalization in addition to the vertex correction.

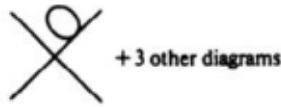
For the renormalization of connected Green's functions, we need to add one-particle reducible diagrams and attach propagators for the external lines. We want to show that the renormalized Green functions when expressed in terms of renormalized quantities are completely finite. We start with the unrenormalized Green's function of the form,

$$G_0^{(4)}(p_1, \dots, p_4) = \Delta^{(0)}(p_j) \left\{ -i\lambda_0 + 3\Gamma(s_0) + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) + \right. \\ \left. + (-i\lambda_0) \sum_{k=1}^4 \left[ -i\Sigma(p_k^2) i\Delta^{(0)}(p_k) \right] \right\} , \quad (11.37)$$

where

$$\Delta^{(0)}(p_j) = \frac{1}{p_j^2 - \mu_0^2 + i\varepsilon} \quad (11.38)$$

is the zeroth order bare propagator and the last terms here are coming from the diagrams of the following type,



1P reducible 4-point function

We can combine the first term and the last terms in  $G_0^{(4)}(p_1, \dots, p_4)$  to get

$$\begin{aligned}
& (-i\lambda_0) \left\{ 1 + \sum_{k=1}^4 \left[ \Sigma(p_k^2) \Delta^{(0)}(p_k) \right] \right\} \simeq \\
& \simeq (-i\lambda_0) \left[ \prod_{k=1}^4 \frac{1}{1 - \Sigma(p_k^2) \Delta^{(0)}(p_k)} \right] + O(\lambda_0^3) = \\
& = (-i\lambda_0) \prod_{k=1}^4 \left\{ \left[ \Delta^{(0)}(p_k) \right]^{-1} \frac{1}{[p_k^2 - \mu_0^2 - \Sigma(p_k^2)]} \right\} = \quad (11.39) \\
& = (-i\lambda_0) \prod_{k=1}^4 \left\{ \left[ \Delta^{(0)}(p_k) \right]^{-1} \Delta(p_k) \right\},
\end{aligned}$$

where

$$\Delta(p_k) = \frac{1}{[p_k^2 - \mu_0^2 - \Sigma(p_k^2)]}. \quad (11.40)$$

Since the difference between  $\Delta(p_k)$  and  $\Delta^{(0)}(p_k)$  is higher order in  $\lambda_0$ , we can make the approximation for the rest of the terms in (11.37),

$$\begin{aligned}
& \Delta^{(0)}(p_j) \left[ 3\Gamma(s_0) + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) \right] \simeq \\
& \simeq \prod_{j=1}^4 \Delta(p_j) \left[ 3\Gamma(s_0) + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) \right]. \quad (11.41)
\end{aligned}$$

The unrenormalized Green's function is then

$$\begin{aligned}
G_0^{(4)}(p_1, \dots, p_4) &= \left[ \prod_{j=1}^4 \Delta(p_j) \right] \left[ -i\lambda_0 + 3\Gamma(s_0) + \tilde{\Gamma}(s) + \tilde{\Gamma}(t) + \tilde{\Gamma}(u) \right] = \\
&= \left[ \prod_{j=1}^4 \Delta(p_j) \right] \Gamma_0^{(4)}(s, t, u). \quad (11.42)
\end{aligned}$$

We now multiply the unrenormalized Green's function by the appropriate factor of  $Z_\varphi$  to get the renormalized one,

$$\begin{aligned}
G_R^{(4)}(p_1, \dots, p_4) &= Z_\varphi^{-2} G_0^{(4)}(p_1, \dots, p_4) = Z_\varphi^{-2} \left[ \prod_{j=1}^4 \Delta(p_j) \right] \Gamma_0^{(4)}(s, t, u) = \\
&= Z_\varphi^{-2} \left[ Z_\varphi^4 \prod_{j=1}^4 i\Delta_R(p_j) \right] Z_\varphi^{-2} \Gamma_R^{(4)}(s, t, u) = \quad (11.43) \\
&= \left[ \prod_{j=1}^4 i\Delta_R(p_j) \right] \Gamma_R^{(4)}(s, t, u).
\end{aligned}$$

Thus we have removed all the divergences in the connected 4-point Green's function.

In summary, Green's functions can be made finite if we express the bare quantities in terms of the renormalized ones through the relations,

$$\varphi = Z_\varphi^{-1/2} \varphi_0, \quad \lambda = Z_\lambda^{-1} Z_\varphi^2 \lambda_0, \quad \mu^2 = \mu_0^2 + \delta\mu^2, \quad (11.44)$$

where  $\delta\mu^2 = \Sigma(\mu^2)$ . For an  $n$ -point Green's function when we express the bare mass  $\mu_0$  and bare coupling  $\lambda_0$  in terms of the renormalized mass  $\mu$  and coupling  $\lambda$ , and multiply by  $Z_\varphi^{-1/2}$  for each external line the result (the renormalized  $n$ -point Green's function) is completely finite,

$$G_R^{(n)}(p_1, \dots, p_n; \lambda, \mu) = Z_\varphi^{-n/2} G_0^{(n)}(p_1, \dots, p_n; \lambda_0, \mu_0, \Lambda), \quad (11.45)$$

where  $\Lambda$  is the cutoff needed to define the divergent integrals. This feature, in which all the divergences, after rewriting  $\mu_0$  and  $\lambda_0$  in terms of  $\mu$  and  $\lambda$  are aggregated into some multiplicative constants, is called being *multiplicatively renormalizable*.

## 11.2 BPH Renormalization

BPH renormalization (after *Bogoliubov, Parasiuk, Hepp*) is completely equivalent to the conventional renormalization but organized differently. We will illustrate this also in the simple  $\lambda\varphi^4$  theory.

Start from the unrenormalized Lagrangian,

$$\mathcal{L}_0 = \frac{1}{2} \left[ (\partial_\mu \varphi_0)^2 - \mu_0^2 \varphi_0^2 \right] - \frac{\lambda_0}{4!} \varphi_0^4, \quad (11.46)$$

where all the quantities are unrenormalized. We can rewrite this in terms of renormalized quantities using (11.44),

$$\mathcal{L}_0 = \mathcal{L} + \Delta\mathcal{L}, \quad (11.47)$$

where

$$\mathcal{L} = \frac{1}{2} \left[ (\partial_\mu \varphi)^2 - \mu^2 \varphi^2 \right] - \frac{\lambda}{4!} \varphi^4 \quad (11.48)$$

has exactly the same form as the original Lagrangian, is called the *renormalized Lagrangian*, and by so called counterterm Lagrangian:

$$\Delta\mathcal{L} = \frac{(Z_\varphi - 1)}{2} \left[ (\partial_\mu \varphi)^2 - \mu^2 \varphi^2 \right] + \frac{\delta\mu^2}{2} Z_\varphi \varphi^2 - \frac{\lambda(Z_\lambda - 1)}{4!} \varphi^4, \quad (11.49)$$

which contains all the divergent constants,  $Z_\varphi$ ,  $Z_\lambda$ ,  $\delta\mu^2$ .

The BPH renormalization scheme consists of the following steps;

1. Start with renormalized Lagrangian given in (11.48) to construct propagators and vertices;
2. Isolate the divergent parts of 1PI diagrams by Taylor expansion. Construct a set of counterterms  $\Delta\mathcal{L}^{(1)}$  which is designed to cancel these one-loop divergences;
3. A new Lagrangian  $\mathcal{L}^{(1)} = \mathcal{L} + \Delta\mathcal{L}^{(1)}$  is used to generate the 2-loop diagrams and to construct the counterterms  $\Delta\mathcal{L}^{(2)}$  which cancels the divergences up to this order and so on, as this sequence of operations is iteratively applied.

The resulting Lagrangian is of the form,

$$\mathcal{L}^{(\infty)} = \mathcal{L} + \Delta\mathcal{L} , \quad (11.50)$$

where the counterterm Lagrangian  $\Delta\mathcal{L}$  is given by,

$$\Delta\mathcal{L} = \Delta\mathcal{L}^{(1)} + \Delta\mathcal{L}^{(2)} + \dots \Delta\mathcal{L}^{(n)} + \dots . \quad (11.51)$$

We will now show that the counter term Lagrangian has the same structure as that in (11.49).

### 11.2.1 Power Counting Method

This method will help to classify divergences systematically. For a given Feynman diagram, we define *superficial degree of divergence*  $D$  as the number of loop momenta in the numerator minus the number of loop momenta in the denominator. For illustration we will compute  $D$  in  $\lambda\phi^4$  theory. Define

$$\begin{aligned} B &= \text{number of external lines,} \\ IB &= \text{number of internal lines,} \\ n &= \text{number of vertices.} \end{aligned} \quad (11.52)$$

It is easy to see that the superficial degree of divergence is given by

$$D = 4 - B . \quad (11.53)$$

It is important to note that  $D$  depends only on the number of external lines,  $B$  and not on  $n$ , the number of vertices. This is a consequence of  $\lambda\phi^4$  theory and might not hold for other interactions. From (11.53) we see that  $D \geq 0$  only for  $B = 2, 4$  ( $B = \text{even}$  because of the symmetry  $\phi \rightarrow -\phi$ ). In the analysis of divergences, we will use the superficial degree of divergences to construct the counterterms. The reason for this will be explained later.

1.  $B = 2, \Rightarrow D = 2$

Being quadratically divergent, the necessary Taylor expansion for the 2-point function is of the form,

$$\Sigma(p^2) = \Sigma(0) + p^2 \Sigma'(0) + \tilde{\Sigma}(p^2) , \quad (11.54)$$

where  $\Sigma(0)$  and  $\Sigma'(0)$  are divergent and  $\tilde{\Sigma}(p^2)$ , To cancel these divergences we need to add two counterterms,

$$\frac{1}{2} \Sigma(0) \phi^2 + \frac{1}{2} \Sigma'(0) (\partial_\mu \phi)^2 , \quad (11.55)$$

which give the following contributions,



2.  $B = 4, \Rightarrow D = 0$

The Taylor expansion is

$$\Gamma^{(4)}(p_i) = \Gamma^{(4)}(0) + \tilde{\Gamma}^{(4)}(p_i) , \quad (11.56)$$

where  $\Gamma^{(4)}(0)$  is logarithmically divergent which is to be cancelled by counterterm of the form

$$\frac{i}{4!} \Gamma^{(4)}(0) \phi^4 \quad (11.57)$$



The general counterterm Lagrangian is then

$$\Delta \mathcal{L} = \frac{1}{2} \Sigma(0) \phi^2 + \frac{1}{2} \Sigma'(0) (\partial_\mu \phi)^2 + \frac{i}{4!} \Gamma^{(4)}(0) \phi^4 , \quad (11.58)$$



which is clearly the same as (11.49) with the identification

$$\begin{aligned}\Sigma'(0) &= (Z_\varphi - 1) , \\ \Sigma(0) &= -(Z_\varphi - 1)\mu^2 + \delta\mu^2 ,\end{aligned}\tag{11.59}$$

$$\Gamma^{(4)}(0) = -i\lambda(1 - Z_\lambda) .\tag{11.60}$$

This illustrates the equivalence of BPH renormalization and conventional renormalization.

### 11.2.2 More on BPH Renormalization

The BPH renormalization scheme looks very simple. It is remarkable that this simple scheme can serve as the basis for setting up a proof for a certain class of field theory. There are many interesting and useful features in BPH which do not show themselves on the first glance and are very useful in the understanding of this renormalization program. We will now discuss some of them.

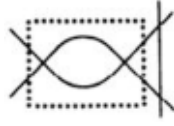
**Convergence of Feynman Diagrams.** In our analysis we have used the superficial degree of divergences  $D$ . To 1-loop order that superficial degree of divergence is the same as the real degree of divergence. When we go beyond 1-loop it is possible to have an overall  $D < 0$  while there are real divergences in the subgraphs. The real convergence of a Feynman graph is governed by Weinberg's theorem: *The general Feynman integral converges if the superficial degree of divergence of the graph together with the superficial degree of divergence of all subgraphs are negative.* To be more explicit, consider a Feynman graph with  $n$  external lines and  $l$  loops. Introduce a cutoff  $\Lambda$  in the momentum integration to estimate the order of divergence,

$$\Gamma^{(n)}(p_1, \dots, p_{n-1}) = \int_0^\Lambda d^4 q_1 \cdots d^4 q_l I(p_1, \dots, p_{n-1}; q_1, \dots, q_l) .\tag{11.61}$$

Take a subset  $S = \{q'_1, q'_2, \dots, q'_m\}$  of the loop momenta  $\{q_1, \dots, q_l\}$  and scale them to infinity and all other momenta fixed. Let  $D(S)$  be the superficial degree of divergence associated with integration over this set, i.e.

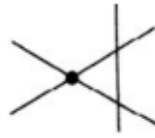
$$\left| \int_0^\Lambda d^4 q'_1 \cdots d^4 q'_m I \right| \leq \Lambda^{D(S)} \{\ln \Lambda\} ,\tag{11.62}$$

where  $\{\ln \Lambda\}$  is some function of  $\ln \Lambda$ . Then the convergent theorem states that the integral over  $\{q_1, \dots, q_l\}$  converges if the  $D(S)$ 's for all possible choice of  $S$  are negative. For example the graph in the following figure



divergence in 6-point function

is a 6-point function with  $D = -2$ . But the integration inside the box with  $D = 0$  is logarithmically divergent. However, in the BPH procedure this subdivergence is removed by lower order counter terms as shown below.



counterterm for 6-point function

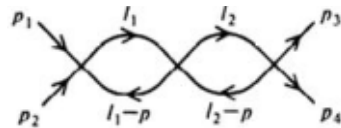
**Classification of Divergent Graphs.** It is useful to distinguish divergent graphs with different topologies in the construction of counterterms.

1. Primitively divergent graphs

A primitively divergent graph has a nonnegative overall superficial degree of divergence but is convergent for all subintegrations. Thus these are diagrams in which the only divergences is caused by all of the loop momenta growing large together. This means that when we differentiate with respect to external momenta at least one of the internal loop momenta will have more power in the denominator and will improve the convergence of the diagram. It is then clear that all the divergences can be isolated in the first few terms of the Taylor expansion.

2. Disjointed divergent graphs

Here the divergent subgraphs are disjointed. For illustration, consider the 2-loop graph given below,



2-loop disjoint divergence

It is clear that differentiating with respect to the external momentum will improve only one of the loop integration but not both. As a result, not all divergences in this diagram can be removed by subtracting out

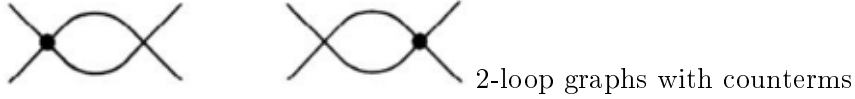
the first few terms in the Taylor expansion around external momenta. However, the lower order counter terms in the BPH scheme will come in to save the day. The Feynman integral is written as

$$\Gamma_a^{(4)}(p) \propto \lambda^3 [\Gamma(p)]^2, \tag{11.63}$$

with

$$\Gamma(p) = \frac{1}{2} \int d^4l \frac{1}{l^2 - \mu^2 + i\varepsilon} \frac{1}{[(l-p)^2 - \mu^2 + i\varepsilon]}, \tag{11.64}$$

and  $p = p_1 + p_2$ . Since  $\Gamma(p)$  is logarithmic divergent,  $\Gamma_a^{(4)}(p)$  cannot be made convergent no matter how many derivatives act on it, even though the overall superficial degree of divergence is zero. However, we have the lower order counterterm  $-\lambda^2\Gamma(0)$  corresponding to the subtraction introduced at the 1-loop level. This generates the additional contributions given in the following diagrams,



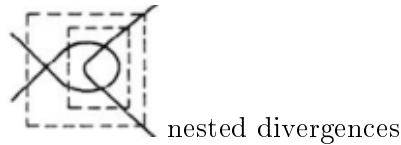
which are proportional to  $-\lambda^3\Gamma(0)\Gamma(p)$ . Adding these 3 contributions, we get

$$\begin{aligned} & \lambda^3 [\Gamma(p)]^2 - 2\lambda^3\Gamma(0)\Gamma(p) = \\ & = \lambda^3 [\Gamma(p) - \Gamma(0)]^2 - \lambda^3 [\Gamma(0)]^2, \end{aligned} \tag{11.65}$$

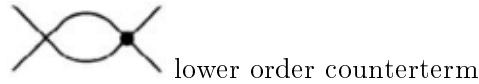
Since the combination in the first  $[\dots]$  is finite, the divergence in the last term can be removed by one differentiation. Here we see that with the inclusion of lower order counterterms, the divergences take the form of polynomials in external momenta. Thus for graphs with disjointed divergences we need to include the lower order counter terms to remove the divergences by subtractions in Taylor expansion.

### 3. Nested divergent graphs

In this case one of a pair of divergent 1PI is entirely contained within the other as shown in the following diagram,



After the subgraph divergence is removed by diagrams with lower order counterterms, the overall divergences is then renormalized by a  $\lambda^3$  counter terms as shown below,



lower order counterterm

Again diagrams with lower-order counterterm insertions must be included in order to aggregate the divergences into the form of polynomial in external momenta.

#### 4. Overlapping divergent graphs

These diagrams are those divergences which are neither nested nor disjointed. These are most difficult to analyze. An example of this is shown below,



overlapping divergences

From these discussion, it is clear that BPH renormalization scheme is quite useful in organizing the higher order divergences in a more systematic way for the removing of divergences by constructing the counterterms.

The general analysis of the renormalization program has been carried out by Bogoliubov, Parasiuk, Hepp. The result is known as BPH theorem, which states: *For a general renormalizable QFT, to any order in perturbation theory, all divergences are removed by the counterterms corresponding to superficially divergent amplitudes.*

### 11.3 Power Counting and Renormalizability

We now discuss the problem of renormalization for more general interactions. It is clear that it is advantageous to use the BPH scheme in this discussion.

### 11.3.1 Theories with Fermions and Scalar Fields

We first study the simple case with fermion  $\psi$  and scalar field  $\phi$ . Write the Lagrangian density as

$$\mathcal{L} = \mathcal{L}_0 + \sum_i \mathcal{L}_i, \quad (11.66)$$

where  $\mathcal{L}_0$  is the free Lagrangian quadratic in the fields and  $\mathcal{L}_i$  are the interaction terms e.g.

$$\mathcal{L}_i = g_1 \bar{\psi} \gamma^\mu \psi \partial_\mu \phi; g_2 (\bar{\psi} \psi)^2; g_3 (\bar{\psi} \psi) \phi, \dots \quad (11.67)$$

Here  $\psi$  denotes a fermion field and  $\phi$  a scalar field. Define the following quantities

$$\begin{aligned} n_i &= \text{number of } i\text{-th type vertices,} \\ b_i &= \text{number of scalar lines in } i\text{-th type vertex,} \\ f_i &= \text{number of fermion lines in } i\text{-th type vertex,} \\ d_i &= \text{number of derivatives in } i\text{-th type of vertex,} \\ B &= \text{number of external scalar lines,} \\ F &= \text{number of external fermion lines,} \\ IB &= \text{number of internal scalar lines,} \\ IF &= \text{number of internal fermion lines.} \end{aligned} \quad (11.68)$$

Counting the scalar and fermion lines, we get

$$B + 2(IB) = \sum_i n_i b_i, \quad (11.69)$$

$$F + 2(IF) = \sum_i n_i f_i, \quad (11.70)$$

Using momentum conservation at each vertex we can compute the number of loop integration  $L$  as

$$L = (IB) + (IF) - n + 1, \quad n = \sum_i n_i, \quad (11.71)$$

where the last term is due to the overall momentum conservation which does not contain the loop integrations. The superficial degree of divergence is then given by

$$D = 4L - 2(IB) - (IF) + \sum_i n_i d_i. \quad (11.72)$$

Using the relations given in (11.69) and (11.70) we get

$$D = 4 - B - \frac{3}{2}F + \sum_i n_i \delta_i, \quad (11.73)$$

where

$$\delta_i = b_i + \frac{3}{2}f_i + d_i - 4 \quad (11.74)$$

is called the *index of divergence* of the interaction. Using the fact that Lagrangian density  $\mathcal{L}$  has dimension 4 and scalar field, fermion field and the derivative have dimensions, 1, 3/2, and 1 respectively, we get for the dimension of the coupling constant  $g_i$  as

$$\dim(g_i) = 4 - b_i - \frac{3}{2}f_i - d_i = -\delta_i. \quad (11.75)$$

We distinguish 3 different situations;

1.  $\delta_i < 0$

In this case,  $D$  decreases with the number of  $i$ -th type of vertices and the interaction is called *super-renormalizable interaction*. The divergences occur only in some lower order diagrams. There is only one type of theory in this category, namely  $\phi^3$  interaction.

2.  $\delta_i = 0$

Here  $D$  is independent of the number of  $i$ -th type of vertices and interactions are called *renormalizable interactions*. The divergence are present in all higher-order diagrams of a finite number of Green's functions. Interactions in this category are of the form,  $g\phi^4$ ,  $f(\bar{\psi}\psi)\phi$ .

3.  $\delta_i > 0$

Then  $D$  increases with the number of the  $i$ -th type of vertices and all Green's functions are divergent for large enough  $n_i$ . These are called *non-renormalizable interactions*. There are plenty of examples in this category,  $g_1\bar{\psi}\gamma^\mu\psi\partial_\mu\phi$ ,  $g_2(\bar{\psi}\psi)^2$ ,  $g_3\phi^5, \dots$ .

The index of divergence  $\delta_i$  can be related to the operator's *canonical dimension* which is defined in terms of the high energy behavior in the free field theory. For any operator  $A$  we write the 2-point function as

$$D_A(p^2) = \int d^4x e^{-ip \cdot x} \langle 0 | T(A(x)A(0)) | 0 \rangle. \quad (11.76)$$

If the asymptotic behavior is of the form,

$$D_A(p^2) \longrightarrow (p^2)^{-\omega_A/2}, \quad \text{as } p^2 \longrightarrow \infty, \quad (11.77)$$

then the canonical dimension is defined as

$$d(A) = \frac{4 - \omega_A}{2}. \quad (11.78)$$

Thus for the case of fermion and scalar fields we have,

$$\begin{aligned} d(\phi) &= 1, & d(\partial^n \phi) &= 1 + n, \\ d(\psi) &= \frac{3}{2}, & d(\partial^n \psi) &= \frac{3}{2} + n. \end{aligned} \quad (11.79)$$

Note that in these simple cases, these values are the same as those obtained in the dimensional analysis in the classical theory and sometimes they are also called the naive dimensions. As we will see later for the vector field, the canonical dimension is not necessarily the same as the naive dimension.

For composite operators that are polynomials in the scalar or fermion fields it is difficult to know their asymptotic behavior. So we define their canonical dimensions as the algebraic sum of their constituent fields. For example,

$$d(\phi^2) = 2, \quad d(\bar{\psi}\psi) = 3. \quad (11.80)$$

For general composite operators that show up in the those interaction described before, we have,

$$d(\mathcal{L}_i) = b_i + \frac{3}{2}f_i + d_i \quad (11.81)$$

and it is related to the index of divergence as

$$\delta_i = d(\mathcal{L}_i) - 4. \quad (11.82)$$

We see that a dimension 4 interaction is renormalizable and greater than 4 is non-renormalizable.

**Counter Terms.** Recall that we add counterterms to cancel all the divergences in Green's functions with superficial degree of divergences  $D \geq 0$ . For convenience we use the Taylor expansion around zero external momenta  $p_i = 0$ . It is easy to see that a general diagram with  $D \geq 0$ , counter terms will be of the form

$$O_{ct} = (\partial_\mu)^\alpha (\psi)^F (\phi)^B, \quad (\alpha = 1, 2, \dots, D) \quad (11.83)$$

and the canonical dimension is

$$d_{ct} = \frac{3}{2}F + B + \alpha . \quad (11.84)$$

The index of divergence of the counterterms is

$$\delta_{ct} = d_{ct} - 4 . \quad (11.85)$$

Using the relation in (11.73) we can write this as

$$\delta_{ct} = (\alpha - D) + \sum_i n_i \delta_i . \quad (11.86)$$

Since  $\alpha \leq D$ , we have the result

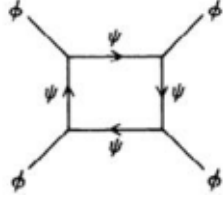
$$\delta_{ct} \leq \sum_i n_i \delta_i . \quad (11.87)$$

Thus, the counterterms induced by a Feynman diagrams have indices of divergences less or equal to the sum of the indices of divergences of all interactions  $\delta_i$  in the diagram.

We then get the general result that the renormalizable interactions which have  $\delta_i = 0$  will generate counterterms with  $\delta_{ct} \leq 0$ . Thus if all the  $\delta_i \leq 0$  terms are present in the original Lagrangian, so that the counter terms have the same structure as the interactions in the original Lagrangian, they may be considered as redefining parameters like masses and coupling constants in the theory. On the other hand non-renormalizable interactions which have  $\delta_i > 0$  will generate counterterms with arbitrary large  $\delta_{ct}$  in sufficiently high orders and clearly cannot be absorbed into the original Lagrangian by a redefinition of parameters  $\delta_{ct}$ . Thus non-renormalizable theories will not necessarily be infinite; however the infinite number of counterterms associated with a non-renormalizable interaction will make it lack in predictive power and hence be unattractive, in the framework of perturbation theory.

We will adopt a more restricted definition of renormalizability: a Lagrangian is said to be renormalizable by power counting if all the counterterms induced by the renormalization procedure can be absorbed by redefinitions of parameters in the Lagrangian. With this definition the theory with Yukawa interaction  $\bar{\psi}\gamma_5\psi\phi$  by itself, is not renormalizable even though the coupling constant is dimensionless. This is because the 1-loop diagram shown below





box diagram for Yukawa coupling

is logarithmically divergent and needs a counter term of the form  $\phi^4$  which is not present in the original Lagrangian. Thus Yukawa interaction with additional  $\phi^4$  interaction is renormalizable.

### 11.3.2 Theories with Vector Fields

Here we distinguish massless from massive vector fields because their asymptotic behaviors for the free field propagators are very different.

Massless vector field is usually associated with local gauge invariance as in the case of QED. The asymptotic behavior of free field propagator for such vector field is very similar to that of scalar field. For example, in the Feynman gauge we have

$$\Delta_{\mu\nu}(k) = \frac{-ig_{\mu\nu}}{k^2 + i\epsilon} \longrightarrow O(k^{-2}), \quad \text{for large } k^2, \quad (11.88)$$

which has the same asymptotic behavior as for scalar fields. Thus the power counting for theories with massless vector field interacting with fermions and scalar fields is the same as before. The renormalizable interactions in this category are of the type,

$$\bar{\psi}\gamma_\mu\psi A^\mu, \quad \phi^2 A_\mu A^\mu, \quad (\partial_\mu\phi)\phi A^\mu. \quad (11.89)$$

Here  $A^\mu$  is a massless vector field and  $\psi$  a fermion field.

The free Lagrangian of massive vector field is of the form,

$$\mathcal{L}_0 = -\frac{1}{4}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2 + \frac{1}{2}M_V^2 V_\mu^2, \quad (11.90)$$

where  $V_\mu$  is a massive vector field and  $M_V$  is the mass of the vector field. The propagator in momentum space is of the form,

$$D_{\mu\nu}(k) = \frac{-i(g_{\mu\nu} - k_\mu k_\nu / M_V^2)}{k^2 - M_V^2 + i\epsilon} \longrightarrow O(1), \quad \text{as } k \rightarrow \infty. \quad (11.91)$$

This implies that canonical dimension of massive vector field is two rather than one. The power counting is now modified with superficial degree of divergence given by

$$D = 4 - B - \frac{3}{2}F - V + \sum_i n_i (\Delta_i - 4) , \quad (11.92)$$

with

$$\Delta_i = b_i + \frac{3}{2}f_i + 2v_i + d_i . \quad (11.93)$$

Here  $V$  is the number of external vector lines,  $v_i$  is the number of vector fields in the  $i$ -th type of vertex and  $\Delta_i$  is the canonical dimension of the interaction term in  $\mathcal{L}$ . From the formula for  $\Delta_i$  we see that the only renormalizable interaction involving massive vector field,  $\Delta_i \leq 4$ , is of the form,  $\phi^2 A_\mu$  and is not Lorentz invariant. Thus there is no nontrivial interaction of the massive vector field which is renormalizable. However, two important exceptions should be noted;

1. In a gauge theory with spontaneous symmetry breaking, the gauge boson will acquire mass in such a way to preserve the renormalizability of the theory;
2. A theory with a neutral massive vector boson coupled to a conserved current is also renormalizable. Heuristically, we can understand this as follows. The propagator in (11.91) always appears between conserved currents  $J^\mu(k)$  and  $J^\nu(k)$  and the  $k_\mu k_\nu / M_V^2$  term will not contribute because of current conservation,  $k^\mu J_\mu(k) = 0$ , or in the coordinate space  $\partial^\mu J_\mu(x) = 0$ . Then the power counting is essentially the same as for the massless vector field case.

### 11.3.3 Renormalization of Composite Operators

In some cases, we need to consider the Green's function of composite operator, an operator with more than one fields at the same space time. Consider a simple composite operator of the form  $\Omega(x) = \frac{1}{2}\phi^2(x)$  in  $\lambda\phi^4$  theory. Green's function with one insertion of  $\Omega$  is of the form,

$$G_\Omega^{(n)}(x; x_1, x_2, x_3, \dots, x_n) = \langle 0 | T \left( \frac{1}{2}\phi^2(x)\phi(x_1)\phi(x_2)\dots\phi(x_n) \right) | 0 \rangle , \quad (11.94)$$

In momentum space we have

$$\begin{aligned} (2\pi)^4 \delta^4(p + p_1 + p_2 + \dots + p_n) G_{\phi^2}^{(n)}(p; p_1, p_2, p_3, \dots, p_n) = \\ = \int d^4x e^{-ipx} \int \prod_{i=1}^n d^4x_i e^{-ip_i x_i} G_{\Omega}^{(n)}(x; x_1, x_2, x_3, \dots, x_n) . \end{aligned} \quad (11.95)$$

In perturbation theory, we can use Wick's theorem to work out these Green's functions in terms of Feynman diagram. Example, to lowest order in  $\lambda$  the 2-point function with one composite operator  $\Omega(x) = \frac{1}{2}\phi^2(x)$  is, after using the Wick's theorem,

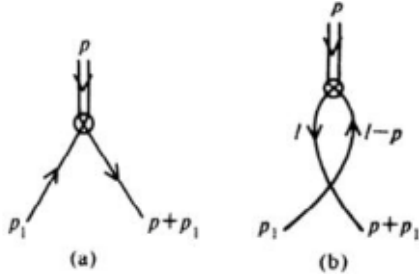
$$\begin{aligned} G_{\phi^2}^{(2)}(x; x_1, x_2) &= \frac{1}{2} \langle 0|T\{\phi^2(x)\phi(x_1)\phi(x_2)\}|0\rangle = \\ &= i\Delta(x - x_1)i\Delta(x - x_2) , \end{aligned} \quad (11.96)$$

or in momentum space

$$G_{\phi^2}^{(2)}(p; p_1, p_2) = i\Delta(p_1)i\Delta(p + p_1) . \quad (11.97)$$

If we truncate the external propagators, we get

$$\Gamma_{\phi^2}^{(2)}(p, p_1, -p_1 - p) = 1 . \quad (11.98)$$



graphs for composite operator

To first order in  $\lambda$ , we have

$$\begin{aligned} G_{\phi^2}^{(2)}(x, x_1, x_2) &= \int \langle 0|T\{\frac{1}{2}\phi^2(x)\phi(x_1)\phi(x_2)\frac{-i\lambda}{4!}\phi^4(y)\}|0\rangle d^4y = \\ &= \int d^4y \frac{-i\lambda}{2} [i\Delta(x - y)]^2 i\Delta(x_1 - y) i\Delta(x_2 - y) . \end{aligned} \quad (11.99)$$

The amputated 1PI momentum space Green's function is

$$\Gamma_{\phi^2}^{(2)}(p; p_1, -p-p_1) = \frac{-i\lambda}{2} \int \frac{d^4l}{(2\pi)^4} \frac{i}{l^2 - \mu^2 + i\epsilon} \frac{i}{(l-p)^2 - \mu^2 + i\epsilon} . \quad (11.100)$$

To calculate this type of Green's functions systematically, we can add a term  $\chi(x)\Omega(x)$  to  $\mathcal{L}$

$$\mathcal{L}[\chi] = \mathcal{L}[0] + \chi(x)\Omega(x) , \quad (11.101)$$

where  $\chi(x)$  is a c-number source function. We can construct the generating functional  $W[\chi]$  in the presence of this external source. We obtain the connected Green's function by differentiating  $\ln W[\chi]$  with respect to  $\chi$  and then setting  $\chi$  to zero.

Superficial degrees of divergence for Green's function with one composite operator is,

$$D_\Omega = D + \delta_\Omega = D + (d_\Omega - 4) , \quad (11.102)$$

where  $d_\Omega$  is the canonical dimension of  $\Omega$ . For the case of  $\Omega(x) = \frac{1}{2}\phi^2(x)$ ,  $d_{\phi^2} = 2$  and  $D_{\phi^2} = 2 - n \Rightarrow$  only  $\Gamma_{\phi^2}^{(2)}$  is divergent. Taylor expansion takes the form,

$$\Gamma_{\phi^2}^{(2)}(p; p_1) = \Gamma_{\phi^2}^{(2)}(0, 0) + \Gamma_{\phi^2 R}^{(2)}(p, p_1) . \quad (11.103)$$

We can combine the counter term,

$$\frac{-i}{2}\Gamma^{(2)}_{\phi^2}(0, 0)\chi(x)\phi^2(x) , \quad (11.104)$$

with the original term to write

$$\frac{-i}{2}\chi\phi - \frac{i}{2}\Gamma_{\phi^2}^2(0, 0)\chi\phi^2 = -\frac{i}{2}Z_{\phi^2}\chi\phi^2 . \quad (11.105)$$

In general, we need to insert a counterterm  $\Delta\Omega$  into the original addition

$$L \rightarrow L + \chi(\Omega + \Delta\Omega) . \quad (11.106)$$

If  $\Delta\Omega = C\Omega$ , as in the case of  $\Omega = \frac{1}{2}\phi^2$ , we have

$$L[\chi] = L[0] + \chi Z_\Omega \Omega = L[0] + \chi \Omega_0 , \quad (11.107)$$

with

$$\Omega_0 = Z_\Omega \Omega = (1 + C)\Omega . \quad (11.108)$$

Such composite operators are said to be multiplicative renormalizable and Green's functions of unrenormalized operator  $\Omega_0$  is related to that of renormalized operator  $\Omega$  by

$$\begin{aligned} G_{\Omega_0}^{(n)}(x; x_1, x_2, \dots x_n) &= \langle 0 | T \{ \Omega_0(x) \phi(x_1) \phi(x_2) \dots \phi(x_n) \} | 0 \rangle = \\ &= Z_\Omega Z_\phi^{n/2} G_{lR}^{(n)}(x; x_1, \dots x_n) . \end{aligned} \quad (11.109)$$

For more general cases,  $\Delta\Omega \neq C\Omega$  and the renormalization may require counterterms proportional to other composite operators.

As an example consider the case with 2 composite operators  $A$  and  $B$ . Denote the counterterms by  $\Delta A$  and  $\Delta B$ . Including the counter terms we can write,

$$L[\chi] = L[0] + \chi_A(A + \Delta A) + \chi_B(B + \Delta B) . \quad (11.110)$$

Often the counterterms  $\Delta A$  and  $\Delta B$  are linear combinations of  $A$  and  $B$ ,

$$\Delta A = C_{AA}A + C_{AB}B , \quad \Delta B = C_{BA}A + C_{BB}B . \quad (11.111)$$

We can write

$$L[\chi] = L[0] + (\chi_A \ \chi_B) \{C\} \begin{pmatrix} A \\ B \end{pmatrix} , \quad (11.112)$$

where

$$\{C\} = \begin{pmatrix} 1 + C_{AA} & C_{AB} \\ C_{BA} & 1 + C_{BB} \end{pmatrix} . \quad (11.113)$$

Diagonalize  $\{C\}$  by bi-unitary transformation

$$U\{C\}V^+ = \begin{pmatrix} Z_{A'} & 0 \\ 0 & Z_{B'} \end{pmatrix} . \quad (11.114)$$

Then

$$L[\chi] = L[0] + Z_{A'}\chi_{A'}A' + Z_{B'}\chi_{B'}B' . \quad (11.115)$$

Also

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = V \begin{pmatrix} A \\ B \end{pmatrix} , \quad (\chi_{A'} \ \chi_{B'}) = (\chi_A \ \chi_B)U , \quad (11.116)$$

and  $A'$ ,  $B'$  are multiplicatively renormalizable.

### 11.3.4 Symmetry and Renormalization

For a theory with global symmetry, we require that the counter terms should also respect the symmetry. For example, consider the Lagrangian given by

$$\mathcal{L} = \frac{1}{2} [(\partial_\mu\phi_1)^2 + (\partial_\mu\phi_2)^2] - \frac{\mu^2}{2} (\phi_1^2 + \phi_2^2) - \frac{\lambda}{4} (\phi_1^2 + \phi_2^2)^2 . \quad (11.117)$$

This Lagrangian has the  $O(2)$  symmetry given below

$$\begin{aligned}\phi_1 &\rightarrow \phi'_1 = \cos \theta \phi_1 + \sin \theta \phi_2 , \\ \phi_2 &\rightarrow \phi'_2 = -\sin \theta \phi_1 + \cos \theta \phi_2 ,\end{aligned}\tag{11.118}$$

The counter terms for this theory should have the same symmetry. For example the mass counter term should be of the form  $\delta\mu^2 (\phi_1^2 + \phi_2^2)$ , i.e. the coefficient of  $\phi_1^2$  counter term should be the same as  $\phi_2^2$  term. Then the only other possible counter terms are of the form,

$$(\partial_\mu \phi_1)^2 + (\partial_\mu \phi_2)^2, \quad (\phi_1^2 + \phi_2^2)^2 .\tag{11.119}$$

For the case the symmetry is slightly broken an interesting feature occurs. We will illustrate this with a simple case where the symmetry breaking as,

$$\mathcal{L}_{SB} = c (\phi_1^2 - \phi_2^2) .\tag{11.120}$$

Since the index of divergence for  $\mathcal{L}_{SB}$  is  $\delta_{SB} = -2$ , the superficial degree of divergence for graphs containing  $\mathcal{L}_{SB}$  is  $D_{SB} = 4 - B_1 - B_2 - 2n_{SB}$ , where  $B_1$  and  $B_2$  are number of external  $\phi_1$  and  $\phi_2$  lines and  $n_{SB}$  is the number of times  $\mathcal{L}_{SB}$  appears in the graph. For the case  $n_{SB} = 1$ , we have  $D_{SB} = 2 - B_1 - B_2$ . This means that  $D_{SB} \geq 0$  only for  $B_1 = 2, B_2 = 0$ , or  $B_1 = 0, B_2 = 2$  and the counter terms we need are  $\phi_1^2$ , and  $\phi_2^2$ . The combination  $\phi_1^2 + \phi_2^2$  can be absorbed in the mass counter term while the other combination  $\phi_1^2 - \phi_2^2$  can be absorbed into  $\mathcal{L}_{SB}$ . This shows the when the symmetry is broken, the counterterms we need will have the property that,  $\delta_{CT} \leq \delta_{SB}$ . Or in terms of operator dimension

$$\dim(\mathcal{L}_{CT}) \leq \dim(\mathcal{L}_{SB}) .\tag{11.121}$$

Thus when  $\dim(\mathcal{L}_{SB}) \leq 3$ , the dimension of counter terms cannot be 4. This situation is usually referred to as soft breaking of the symmetry. This is known as the *Szymanzik theorem*. Note that for the soft breaking the coupling constant  $g_{SB}$  will have positive dimension of mass and will be negligible when energies become much larger than  $g_{SB}$ . In other words, the symmetry will be restored at high energies.

### 11.3.5 Ward Identity

In case of global symmetry, we also have some useful relation for composite operator like the current operator which generates the symmetry. We will

give a simple illustration of this feature. The Lagrangian given in (11.117) can be rewritten as

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial_\mu \phi - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad (11.122)$$

where  $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ . The symmetry transformation is then  $\phi \rightarrow \phi' = e^{i\theta} \phi$ . This will give rise, through Noether's theorem, the current of the form,

$$J_\mu = i \left[ (\partial_\mu \phi^\dagger) \phi - (\partial_\mu \phi) \phi^\dagger \right] \quad (11.123)$$

is conserved,  $\partial^\mu J_\mu = 0$ . From the canonical commutation relation,

$$\left[ \partial_0 \phi^\dagger(\vec{x}, t), \phi(\vec{x}', t) \right] = -i \delta^3(\vec{x} - \vec{x}'), \quad (11.124)$$

we can derive,

$$\left[ J_0(\vec{x}, t), \phi(\vec{x}', t) \right] = \delta^3(\vec{x} - \vec{x}') \phi(\vec{x}', t), \quad (11.125)$$

$$\left[ J_0(\vec{x}, t), \phi^\dagger(\vec{x}', t) \right] = -\delta^3(\vec{x} - \vec{x}') \phi^\dagger(\vec{x}', t). \quad (11.126)$$

Now consider the Green's function of the form,

$$G_\mu(p, q) = \int d^4x d^4y e^{-iq \cdot x - ip \cdot y} \langle 0 | T \left( J_\mu(x) \phi(y) \phi^\dagger(0) \right) | 0 \rangle. \quad (11.127)$$

Multiply  $q^\mu$  into this Green's function,

$$\begin{aligned} q^\mu G_\mu(p, q) &= -i \int d^4x d^4y e^{-iq \cdot x - ip \cdot y} \partial_x^\mu \langle 0 | T \left( J_\mu(x) \phi(y) \phi^\dagger(0) \right) | 0 \rangle = \\ &= -i \int d^4x e^{-i(q+p) \cdot x} \langle 0 | T \left( \phi(x) \phi^\dagger(0) \right) | 0 \rangle + \\ &\quad + i \int d^4x e^{-ip \cdot y} \langle 0 | T \left( \phi(y) \phi^\dagger(0) \right) | 0 \rangle, \end{aligned} \quad (11.128)$$

where we have used the current conservation and commutators in (11.125) and (11.126). The right-hand side here is just the propagator for the scalar field,

$$\Delta(p) = \int d^4x e^{-ip \cdot x} \langle 0 | T \left( \phi(x) \phi^\dagger(0) \right) | 0 \rangle, \quad (11.129)$$

and we get

$$-iq^\mu G_\mu(p, q) = \Delta(p+q) - \Delta(p). \quad (11.130)$$

This is example of Ward identity.

This relation is derived in terms of unrenormalized fields which satisfy the canonical commutation relation. In terms of renormalized quantities,

$$G_\mu^R(p, q) = Z_\phi^{-1} Z_J^{-1} G_\mu(p, q), \quad \Delta^R(p) = Z_\phi^{-1} \Delta(p), \quad (11.131)$$

the Ward identity in (11.130) becomes

$$-iZ_J q^\mu G_\mu^R(p, q) = \Delta^R(p+q) - \Delta^R(p). \quad (11.132)$$

Since the right-hand side is cutoff independent,  $Z_J$  on the left-hand side is also cutoff independent, and we do not need any counter terms to renormalize  $J_\mu(x)$ . In other words, the conserved current  $J_\mu(x)$  is not renormalized as composite operator, i.e.  $Z_J = 1$ . Thus the relation for the renormalized quantities takes the simple form,

$$-iq^\mu G_\mu^R(p, q) = \Delta^R(p+q) - \Delta^R(p). \quad (11.133)$$

Such a non-renormalization result holds for many conserved quantities.

**Exercise 11.1:** Use the power-counting argument to construct counter-terms for QED.

**Exercise 11.2:** Consider theories with scalar, fermion and massless gauge fields in a  $d$ -dimensional space-time ( $d-1$  space coordinates and 1 time coordinate). Express the superficial degree of divergence of a diagram in terms of the number of external fermion and boson lines.

**Exercise 11.3:** Consider massless two-dimensional QED (the Schwinger model). (i) Calculate the vacuum polarization at one-loop; (ii) Find the full photon propagator and read off the mass of the photon.



## Chapter 12

# Renormalization Group

*The Renormalization Group* (RG) refers to a mathematical apparatus that allows systematic investigation of the changes of a physical system as viewed at different scales. In particle physics, it reflects the changes in the underlying force laws (codified in a quantum field theory), as the energy scale at which physical processes occur varies, energy/momentum and resolution distance scales being effectively conjugate under the uncertainty principle.

A change in scale is called a *scale transformation*. The RG is intimately related to scale invariance and conformal invariance, symmetries in which a system appears the same at all scales (so-called self-similarity).

Variation of the scale is similar to changing of the magnifying power of a 'microscope' viewing the system. In so-called renormalizable theories, the system at one scale will generally be seen to consist of self-similar copies of itself when viewed at a smaller scale, with different parameters describing the components of the system. The components, or fundamental variables, may relate to atoms, elementary particles, atomic spins, etc. The parameters of the theory typically describe the interactions of the components. These may be variable couplings which measure the strength of various forces, or mass parameters. The components themselves may appear to be composed of more of the self-same components as one goes to shorter distances.

For example, in QED an electron appears to be composed of electrons, positrons (anti-electrons) and photons, as one views it at higher resolution, at very short distances. The electron at such short distances has a slightly

different electric charge than does the dressed electron seen at large distances, and this change, or running, in the value of the electric charge is determined by the RG equation.

## 12.1 History of RG

The idea of scale transformations and scale invariance is old in physics. Scaling arguments were commonplace for the Pythagorean School, Euclid and up to Galileo. They became popular again at the end of the XIX century, perhaps the first example being the idea of enhanced viscosity of Reynolds, as a way to explain turbulence.

The RG was initially devised in particle physics, but nowadays its applications extend to solid-state physics, fluid mechanics, cosmology and even nano-technology. An early article by Stueckelberg and Petermann (1953) anticipates the idea in quantum field theory and opened the field conceptually. They noted that renormalization exhibits a group of transformations which transfer quantities from the bare terms to the counter terms. They introduced a function  $h(e)$  in QED, which is now called the beta function.

Gell-Mann and Low in 1954 restricted the idea to scale transformations in QED and focused on asymptotic forms of the photon propagator at high energies. They determined the variation of the electromagnetic coupling in QED, by appreciating the simplicity of the scaling structure of that theory. They thus discovered that the coupling parameter  $g(\mu)$  at the energy scale  $\mu$  is effectively given by the group equation

$$g(\mu) = G^{-1} \left\{ \left( \frac{\mu}{M} \right)^d G[g(M)] \right\} , \quad (12.1)$$

for some function  $G$  (nowadays called Wegner's scaling function) and a constant  $d$ , in terms of the coupling  $g(M)$  at a reference scale  $M$ .

Gell-Mann and Low realized in these results that the effective scale  $\mu$  is arbitrary, and can vary to define the theory at any other scale  $\kappa$ :

$$g(\kappa) = G^{-1} \left\{ \left( \frac{\kappa}{\mu} \right)^d G[g(\mu)] \right\} = G^{-1} \left\{ \left( \frac{\kappa}{M} \right)^d G[g(M)] \right\} . \quad (12.2)$$

The gist of the RG is the property: *As the scale  $\mu$  varies, the theory presents a self similar replica of itself, and any scale can be accessed similarly from*

any other scale, by group action, a formal transitive conjugacy of couplings in the mathematical sense (Schröder's equation).

On the basis of this (finite) group equation and its scaling property, Gell-Mann and Low could then focus on infinitesimal transformations, and invented a computational method based on a mathematical flow function  $\psi(g)$  of the coupling parameter  $g$ , which they introduced. Like the function  $h(e)$  of Stueckelberg and Petermann, their function determines the differential change of the coupling  $g(\mu)$  with respect to a small change in energy scale  $\mu$  through a differential equation, the renormalization group equation:

$$\psi(g) \equiv \beta(g) = \frac{\partial g}{\partial \ln \mu} . \quad (12.3)$$

The modern name is also indicated, *the beta function*, introduced by Callan and Symanzik in 1970. The renormalization group prediction was confirmed 40 years later at the LEP accelerator experiments: the fine structure "constant" of QED was measured to be about  $1/127$  at energies close to 200 GeV, as opposed to the standard low-energy physics value of  $1/137$ .

The RG emerges from the renormalization of the quantum field variables, which normally has to address the problem of infinities in a QFT (although the RG exists independently of the infinities). This problem of systematically handling the infinities of QFT to obtain finite physical quantities was solved for QED by Feynman, Schwinger and Tomonaga, who received the 1965 Nobel Prize for these contributions. They effectively devised the theory of mass and charge renormalization, in which the infinity in the momentum scale is cut off by an ultra-large regulator,  $\Lambda$  (which could ultimately be taken to be infinite). The dependence of physical quantities, such as the electric charge or electron mass, on the scale  $\Lambda$  is hidden, effectively swapped for the longer-distance scales at which the physical quantities are measured. As a result, all observable quantities end up being finite instead, even for an infinite  $\Lambda$ . Gell-Mann and Low realized in RG results that, while infinitesimally a tiny change in  $g$  is provided by the above RG equation given  $\beta(g)$ , the self-similarity is expressed by the fact that  $\beta(g)$  depends explicitly only upon the parameter(s) of the theory, and not upon the scale  $\mu$ . Consequently, the RG group equation (12.3) may be solved for  $g(\mu)$ .

Understanding of the physical meaning and generalization of the renormalizations, which goes beyond the dilation group of conventional renormalizable theories, considers methods where widely different scales of lengths appear simultaneously. It came from condensed matter physics: Kadanoff's paper

in 1966 proposed the "block-spin" RG. The blocking idea is a way to define the components of the theory at large distances as aggregates of components at shorter distances.

This approach covered the conceptual point and was given full computational substance in the extensive important contributions of Wilson. The power of Wilson's ideas was demonstrated by a constructive iterative renormalization solution of a long-standing problem, the Kondo problem, in 1975, as well as the preceding seminal developments of his new method in the theory of second-order phase transitions and critical phenomena in 1971. He was awarded the Nobel prize for these decisive contributions in 1982.

Meanwhile, the RG in particle physics had been reformulated in more practical terms by Callan and Symanzik in 1970. The beta function (12.3), which describes the "running of the coupling" parameter with scale, was also found to amount to the "canonical trace anomaly", which represents the quantum-mechanical breaking of scale (dilation) symmetry in a field theory (Remarkably, quantum mechanics itself can induce mass through the trace anomaly and the running coupling). Applications of the RG to particle physics exploded in number in the 1970s with the establishment of the SM.

In 1973, it was discovered that a theory of interacting colored quarks, called QCD, had a negative beta function. This means that an initial high-energy value of the coupling will eventuate a special value of  $\mu$  at which the coupling blows up (diverges). This special value is the scale of the strong interactions,  $\mu = \Lambda_{QCD}$  and occurs at about 200 MeV. Conversely, the coupling becomes weak at very high energies (asymptotic freedom), and the quarks become observable as point-like particles, in deep inelastic scattering, as anticipated by Feynman-Bjorken scaling. QCD was thereby established as the quantum field theory controlling the strong interactions of particles.

Momentum space RG also became a highly developed tool in solid state physics, but its success was hindered by the extensive use of perturbation theory, which prevented the theory from reaching success in strongly correlated systems. In order to study these strongly correlated systems, variational approaches are a better alternative.

The conformal symmetry is associated with the vanishing of the beta function. This can occur naturally if a coupling constant is attracted, by running, toward a fixed point at which  $\beta(g) = 0$ . In QCD, the fixed point occurs at short distances where  $g \rightarrow 0$  and is called a (trivial) ultraviolet fixed point.

For heavy quarks, such as the top quark, it is calculated that the coupling to the mass-giving Higgs boson runs toward a fixed non-zero (non-trivial) infrared fixed point.

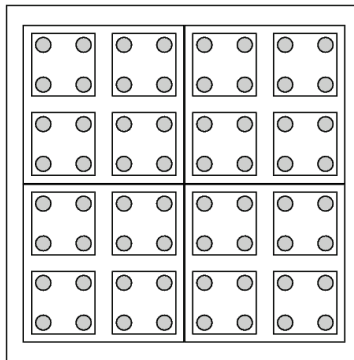
In string theory conformal invariance of the string world-sheet is a fundamental symmetry:  $\beta = 0$  is a requirement. Here,  $\beta$  is a function of the geometry of the space-time in which the string moves. This determines the space-time dimensionality of the string theory and enforces Einstein's equations of general relativity on the geometry.

The RG has become one of the most important tools of modern physics. It is of fundamental importance to string theory and theories of grand unification. Also serves as the modern key idea underlying critical phenomena in condensed matter physics. The RG is often used in combination with the Monte Carlo method.

## 12.2 Block Spin RG

Let us mention the *Block Spin RG* model, which was devised by Kadanoff in 1966.

Consider a 2D solid, a set of atoms in a perfect square array. Assume that atoms interact among themselves only with their nearest neighbors, and that the system is at a given temperature  $T$ . The strength of their interaction is quantified by a certain coupling  $J$ . The physics of the system will be described by a certain formula, say the Hamiltonian  $H(T, J)$ .



Now proceed to divide the solid into blocks of  $2 \times 2$  squares; we attempt to describe the system in terms of block variables, i.e., variables which describe the average behavior of the block. Further assume that, by some lucky coincidence, the physics of block variables is described by a formula of the same kind, but with different values for  $T$  and  $J$ :  $H(T', J')$ . This isn't exactly true, in general, but it is often a good first approximation.

Perhaps, the initial problem was too hard to solve, since there were too many atoms. Now, in the renormalized problem we have only one fourth of them. But why stop now? Another iteration of the same kind leads to  $H(T'', J'')$ , and only one sixteenth of the atoms. We are increasing the observation scale with each RG step.

Of course, the best idea is to iterate until there is only one very big block. Since the number of atoms in any real sample of material is very large, this is more or less equivalent to finding the long range behaviour of the RG transformation which took  $(T, J) \rightarrow (T', J')$  and  $(T', J') \rightarrow (T'', J'')$ . Often, when iterated many times, this RG transformation leads to a certain number of fixed points.

To be more concrete, consider a magnetic system (e.g., the Ising model), in which the  $J$  coupling denotes the trend of neighbor spins to be parallel. The configuration of the system is the result of the tradeoff between the ordering  $J$  term and the disordering effect of temperature.

For many models of this kind there are three fixed points:

1.  $T = 0$  and  $J \rightarrow \infty$ . This means that, at the largest size, temperature becomes unimportant, i.e., the disordering factor vanishes. Thus, in large scales, the system appears to be ordered. We are in a ferromagnetic phase.
2.  $T \rightarrow \infty$  and  $J \rightarrow 0$ . Exactly the opposite; here, temperature dominates, and the system is disordered at large scales.
3. A nontrivial point between them,  $T = T_c$  and  $J = J_c$ . In this point, changing the scale does not change the physics, because the system is in a fractal state. It corresponds to the Curie phase transition, and is also called a critical point.

So, if we are given a certain material with given values of  $T$  and  $J$ , all we

have to do in order to find out the large-scale behaviour of the system is to iterate the pair until we find the corresponding fixed point.

## 12.3 Elementary Considerations

In more technical terms, let us assume that we have a theory described by a certain function  $Z$  of the state variables  $\{s_i\}$  and a certain set of coupling constants  $\{J_k\}$ . This function may be a partition function, an action, a Hamiltonian, etc. It must contain the whole description of the physics of the system.

Now we consider a certain blocking transformation of the state variables  $\{s_i\} \rightarrow \{\tilde{s}_i\}$ , the number of  $\tilde{s}_i$  must be lower than the number of  $s_i$ . Now let us try to rewrite the  $Z$  function only in terms of the  $\tilde{s}_i$ . If this is achievable by a certain change in the parameters,  $\{J_k\} \rightarrow \{\tilde{J}_k\}$ , then the theory is said to be renormalizable.

For some reason, most fundamental theories of physics such as QED, QCD and electro-weak interaction, but not gravity, are exactly renormalizable. Also, most theories in condensed matter physics are approximately renormalizable, from superconductivity to fluid turbulence.

The change in the parameters is implemented by a certain beta function,

$$\{\tilde{J}_k\} = \beta\{J_k\}, \quad (12.4)$$

which is said to induce a renormalization flow (or RG flow) on the  $J$ -space. The values of  $J$  under the flow are called running couplings.

As was stated in the previous section, the most important information in the RG flow are its fixed points. The possible macroscopic states of the system, at a large scale, are given by this set of fixed points. If these fixed points correspond to a free field theory, the theory is said to exhibit quantum triviality, possessing what is called a Landau pole, as in quantum electrodynamics. For a  $\phi^4$  interaction, Aizenman proved that this theory is indeed trivial, for space-time dimension  $D \geq 5$ . For  $D = 4$ , the triviality has yet to be proven rigorously, but lattice computations have provided strong evidence for this. This fact is important as quantum triviality can be used to bound or even predict parameters such as the Higgs boson mass in asymptotic safety scenarios. Numerous fixed points appear in the study of lattice Higgs theories,

but the nature of the quantum field theories associated with these remains an open question.

Since the RG transformations in such systems are lossy (i.e. the number of variables decreases), there need not be an inverse for a given RG transformation. Thus, in such lossy systems, the RG is, in fact, a semigroup.

## 12.4 Relevant Operators and Universality Classes

Consider a certain observable  $A$  of a physical system undergoing an RG transformation. The magnitude of the observable as the length scale of the system goes from small to large may be: (a) always increasing, (b) always decreasing or (c) other. In the first case, the observable is said to be a relevant observable; in the second, irrelevant and in the third, marginal.

A relevant observable is needed to describe the macroscopic behavior of the system; an irrelevant observable is not. Marginal observables may or may not need be taken into account. A remarkable broad fact is that most observables are irrelevant, i.e. the macroscopic physics is dominated by only a few observables in most systems. As an example, in microscopic physics, to describe a system consisting of a mole of  $^{12}\text{C}$  atoms we need of the order of  $10^{23}$  (Avogadro's number) variables, while to describe it as a macroscopic system (12 grams of  $^{12}\text{C}$ ) we only need a few.

Before Wilson's RG approach, there was an astonishing empirical fact to explain: the coincidence of the critical exponents (i.e. the exponents of the reduced-temperature dependence of several quantities near a second order phase transition) in very disparate phenomena, such as magnetic systems, superfluid transition (Lambda transition), alloy physics, etc. Thus, in general, thermodynamic features of a system near a phase transition depend only on a small number of variables, such as dimensionality and symmetry, but are insensitive to details of the underlying microscopic properties of the system.

This coincidence of critical exponents for ostensibly quite different physical systems is called *universality* and is now successfully explained by the RG: essentially by showing that the differences among all such phenomena are, in fact, traceable to such irrelevant observables, while the relevant observables are shared in common.



Thus, many macroscopic phenomena may be grouped into a small set of *universality classes*, specified by the shared sets of relevant observables.

## 12.5 Exact RG Equations

An Exact Renormalization Group Equation is one that takes irrelevant couplings into account. There are several formulations.

1. *The Wilson Exact Renormalization Group Equation* is the simplest conceptually, but is practically impossible to implement. Fourier transform into momentum space after Wick rotating into Euclidean space, insist upon a hard momentum cutoff,  $p^2 \leq \Lambda^2$ , so that the only degrees of freedom are those with momenta less than  $\Lambda$ . The partition function is

$$Z = \int_{p^2 \leq \Lambda^2} \mathcal{D}\phi e^{-S_\Lambda[\phi]} . \quad (12.5)$$

For any positive  $\Lambda'$  (less than  $\Lambda$ ) define  $S_{\Lambda'}$  (a functional over field configurations  $\phi$  whose Fourier transform has momentum support within  $p^2 \leq \Lambda'^2$ ) as

$$e^{-S_{\Lambda'}[\phi]} \stackrel{\text{def}}{=} \int_{\Lambda' \leq p \leq \Lambda} \mathcal{D}\phi e^{-S_\Lambda[\phi]} . \quad (12.6)$$

Obviously,

$$Z = \int_{p^2 \leq \Lambda'^2} \mathcal{D}\phi e^{-S_{\Lambda'}[\phi]} . \quad (12.7)$$

In fact, this transformation is transitive. If you compute  $S_{\Lambda'}$  from  $S_\Lambda$  and then compute  $S_{\Lambda''}$  from  $S_{\Lambda'}$ , this gives you the same Wilsonian action as computing  $S_{\Lambda''}$  directly from  $S_\Lambda$ .

2. *The Polchinski Exact RG Equation* involves a smooth UV regulator cutoff. Basically, the idea is an improvement over the Wilson's case. Instead of a sharp momentum cutoff, it uses a smooth cutoff. Essentially, we suppress contributions from momenta greater than  $\Lambda$  heavily. The smoothness of the cutoff, however, allows us to derive a functional differential equation in the cutoff scale  $\Lambda$ . As in Wilson's approach, we have a different action functional for each cutoff energy scale  $\Lambda$ . Each of these actions are supposed to describe exactly the same model which means that their partition functionals have to match exactly.

In other words, for a real scalar field (generalizations to other fields are obvious),

$$Z_\Lambda[J] = \int \mathcal{D}\phi e^{-S_\Lambda[\phi]+J\cdot\phi} = \int \mathcal{D}\phi e^{-\frac{1}{2}\phi\cdot R_\Lambda\cdot\phi - S_{\text{int}\Lambda}[\phi]+J\cdot\phi}, \quad (12.8)$$

and  $Z_\Lambda$  is really independent of  $\Lambda$ . We have used the condensed de Witt notation here. We have also split the bare action  $S_\Lambda$  into a quadratic kinetic part and an interacting part  $S_{\text{int}\Lambda}$ . This split most certainly isn't clean. The "interacting" part can very well also contain quadratic kinetic terms. In fact, if there is any wave function renormalization, it most certainly will. This can be somewhat reduced by introducing field rescalings.  $R_\Lambda$  is a function of the momentum  $p$  and the second term in the exponent is

$$\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} \tilde{\phi}^*(p) R_\Lambda(p) \tilde{\phi}(p) \quad (12.9)$$

when expanded.

When  $p \ll \Lambda$ ,  $R_\Lambda(p)/p^2$  is essentially 1. When  $p \gg \Lambda$ ,  $R_\Lambda(p)/p^2$  becomes very huge and approaches infinity.  $R_\Lambda(p)/p^2$  is always greater than or equal to 1 and is smooth. Basically, this leaves the fluctuations with momenta less than the cutoff  $\Lambda$  unaffected but heavily suppresses contributions from fluctuations with momenta greater than the cutoff. This is obviously a huge improvement over Wilson.

The condition that

$$\frac{d}{d\Lambda} Z_\Lambda = 0 \quad (12.10)$$

can be satisfied by (but not only by)

$$\begin{aligned} \frac{d}{d\Lambda} S_{\text{int}\Lambda} &= \frac{1}{2} \frac{\delta S_{\text{int}\Lambda}}{\delta\phi} \cdot \left( \frac{d}{d\Lambda} R_\Lambda^{-1} \right) \cdot \frac{\delta S_{\text{int}\Lambda}}{\delta\phi} - \\ &- \frac{1}{2} \text{Tr} \left[ \frac{\delta^2 S_{\text{int}\Lambda}}{\delta\phi\delta\phi} \cdot R_\Lambda^{-1} \right]. \end{aligned} \quad (12.11)$$

Distler claimed without proof that this Exact Renormalization Group Equation is not correct nonperturbatively.

3. *The Effective Average Action Exact Renormalization Group Equation* involves a smooth IR regulator cutoff. The idea is to take all fluctuations right up to an IR scale  $k$  into account. The effective average

action will be accurate for fluctuations with momenta larger than  $k$ . As the parameter  $k$  is lowered, the effective average action approaches the effective action which includes all quantum and classical fluctuations. In contrast, for large  $k$  the effective average action is close to the "bare action". So, the effective average action interpolates between the "bare action" and the effective action.

For a real scalar field, one adds an IR cutoff

$$\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} \tilde{\phi}^*(p) R_k(p) \tilde{\phi}(p) \quad (12.12)$$

to the action  $S$ , where  $R_k$  is a function of both  $k$  and  $p$  such that for  $p \gg k$ ,  $R_k(p)$  is very tiny and approaches 0 and for  $p \ll k$ ,  $R_k(p) \gtrsim k^2$ .  $R_k$  is both smooth and nonnegative. Its large value for small momenta leads to a suppression of their contribution to the partition function which is effectively the same thing as neglecting large-scale fluctuations.

One can use the condensed de Witt notation

$$\frac{1}{2} \phi \cdot R_k \cdot \phi \quad (12.13)$$

for this IR regulator. So,

$$e^{W_k[J]} = Z_k[J] = \int \mathcal{D}\phi e^{-S[\phi] - \frac{1}{2} \phi \cdot R_k \cdot \phi + J \cdot \phi}, \quad (12.14)$$

where  $J$  is the source field. The Legendre transform of  $W_k$  ordinarily gives the effective action. However, the action that we started off with is really  $S[\phi] + \phi \cdot R_k \cdot \phi/2$  and so, to get the effective average action, we subtract off  $\phi \cdot R_k \cdot \phi/2$ . In other words,

$$\phi[J; k] = \frac{\delta W_k}{\delta J}[J] \quad (12.15)$$

can be inverted to give  $J_k[\phi]$  and we define the effective average action  $\Gamma_k$  as

$$\Gamma_k[\phi] \stackrel{\text{def}}{=} \{-W[J_k[\phi]] + J_k[\phi] \cdot \phi\} - \frac{1}{2} \phi \cdot R_k \cdot \phi. \quad (12.16)$$

Hence,

$$\begin{aligned}
\frac{d\Gamma_k[\phi]}{dk} &= -\frac{dW_k[J_k]}{dk} - \frac{\delta W_k}{\delta J} \cdot \frac{dJ_k}{dk} + \frac{dJ_k}{dk} \cdot \phi - \frac{1}{2} \phi \cdot \frac{dR_k}{dk} \cdot \phi = \\
&= -\frac{d}{dk} W_k[J_k[\phi]] - \frac{1}{2} \phi \cdot \frac{d}{dk} R_k \cdot \phi = \\
&= \frac{1}{2} \left\langle \phi \cdot \frac{d}{dk} R_k \cdot \phi \right\rangle_{J_k[\phi]; k} - \frac{1}{2} \phi \cdot \frac{d}{dk} R_k \cdot \phi = \quad (12.17) \\
&= \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta J_k}{\delta \phi} \right)^{-1} \cdot \frac{d}{dk} R_k \right] = \\
&= \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta^2 \Gamma_k}{\delta \phi \delta \phi} + R_k \right)^{-1} \cdot \frac{d}{dk} R_k \right],
\end{aligned}$$

thus

$$\frac{d}{dk} \Gamma_k[\phi] = \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta^2 \Gamma_k}{\delta \phi \delta \phi} + R_k \right)^{-1} \cdot \frac{d}{dk} R_k \right] \quad (12.18)$$

is the Exact Renormalization Group Equation which is also known as the Wetterich equation. As shown by Morris the effective action  $\Gamma_k$  is in fact simply related to Polchinski's effective action  $S_{int}$  via a Legendre transform relation.

As there are infinitely many choices of  $R_k$ , there are also infinitely many different interpolating Exact Renormalization Group Equations. Generalization to other fields, like spinorial fields, is straightforward.

Although Polchinski's and the Effective Average Action approaches look similar, they are based upon very different philosophies. In the Effective Average Action case, the bare action is left unchanged (and the UV cutoff scale is also left unchanged) but the IR contributions to the effective action are suppressed whereas in the Polchinski Exact Renormalization Group Equation, the QFT is fixed once and for all but the "bare action" is varied at different energy scales to reproduce the prespecified model. Polchinski's version is certainly much closer to Wilson's idea in spirit. Note that one uses "bare actions" whereas the other uses effective (average) actions.

**Exercise 12.1:** Derive the RG equation for  $\phi^4$  theory.

**Exercise 12.2:** Find  $\beta$ -function for  $\phi^4$  theory.

**Exercise 12.3:** Find  $\beta$ -function for QED.

## Part V

# Lecture – Introduction to Group Theory



## Chapter 13

# Mathematical Descriptions of Groups

Group theory is, in short, the mathematics of symmetry and we will explain how it relates to physics. The most foundational idea here is: *What doesn't change when something else changes* and a group is a precise and well-defined way of specifying the things that change.

To begin with, we define a *Group*  $(G, \star)$  as a set of *objects* (denoted  $G$ ) and some *operation* on those objects (denoted  $\star$ ) subject to the following:

1.  $\forall g_1, g_2 \in G$ , also  $g_1 \star g_2 \in G$  (*closure*);
2.  $\forall g_1, g_2, g_3 \in G$ , then  $(g_1 \star g_2) \star g_3 = g_1 \star (g_2 \star g_3)$  (*associativity*);
3.  $\exists g \in G$ , denoted  $e$ ,  $\ni \forall g_i \in G$ ,  $e \star g_i = g_i \star e = g_i$  (*identity*);
4.  $\forall g \in G$ ,  $\exists h \in G \ni h \star g = g \star h = e$ , (so  $h = g^{-1}$ ) (*inverse*).

Note that the definition of a group doesn't demand that  $g_i \star g_j = g_j \star g_i$ . This is a very important point, but we will discuss it in more detail later.

Now we explain what this means. By 'objects' we literally mean anything. We could be talking about  $\mathbb{Z}$  or  $\mathbb{R}$ , or we could be talking about a set of Easter eggs all painted different colors.

The meaning of 'some operation', which we are calling  $\star$ , can literally be anything you can do to those objects. A formal definition of what  $\star$  means could be given, but it will be easier to understand with examples.

**Example 1:**  $(G, \star) = (\mathbb{Z}, +)$

Consider the set  $G$  to be  $\mathbb{Z}$ , and the operation to be  $\star = +$ , or simply addition.

We first check closure. If you take any two elements of  $\mathbb{Z}$  and add them together, is the result in  $\mathbb{Z}$ ? In other words, if  $a, b \in \mathbb{Z}$ , is  $a + b \in \mathbb{Z}$ ? Obviously the answer is yes; the sum of two integers is an integer, so closure is met.

Now we check associativity. If  $a, b, c \in \mathbb{Z}$ , it is trivially true that  $a + (b + c) = (a + b) + c$ . So, associativity is met.

Now we check identity. Is there an element  $e \in \mathbb{Z}$  such that when you add  $e$  to any other integer, you get that same integer? Clearly the integer 0 satisfies this. So, identity is met.

Finally, is there an inverse? For any integer  $a \in \mathbb{Z}$ , will there be another integer  $b \in \mathbb{Z}$  such that  $a + b = e = 0$ ? Again, this is obvious,  $a^{-1} = -a$  in this case. So, inverse is met.

So,  $(G, \star) = (\mathbb{Z}, +)$  is a group.

**Example 2:**  $(G, \star) = (\mathbb{R}, +)$

Obviously, any two real numbers added together is also a real number.

Associativity will hold (of course).

The identity is again 0.

And finally, once again,  $-a$  will be the inverse of any  $a \in \mathbb{R}$ .

**Example 3:**  $(G, \star) = (\mathbb{R}, \cdot)$  (multiplication)

Closer is met; two real numbers multiplied together give a real number.

Associativity obviously holds.

Identity also holds. Any real number  $a \in \mathbb{R}$ , when multiplied by 1 is  $a$ .



Inverse, on the other hand, is trickier. For any real number, is there another real number you can multiply by it to get 1? The instinctive choice is  $a^{-1} = 1/a$ . But, this doesn't quite work because of  $a = 0$ . This is the only exception, but because there's an exception,  $(\mathbb{R}, \cdot)$  is not a group.

Note that If we take the set  $\mathbb{R} - \{0\}$  instead of  $\mathbb{R}$ , then  $(\mathbb{R} - \{0\}, \cdot)$  is a group.

**Example 4:**  $(G, \star) = (\{1\}, \cdot)$

This is the set with only the element 1, and the operation is normal multiplication. This is indeed a group, but it is extremely uninteresting, and is called the *Trivial Group*.

**Example 5:**  $(G, \star) = (\mathbb{Z}_3, +)$

This is the set of integers mod 3, containing only the elements 0, 1, and 2 (3 mod 3 is 0, 4 mod 3 is 1, 5 mod 3 is 2, etc.)

You can check yourself that this is a group.

## 13.1 Finite Discrete Groups

From the examples above, several things should be apparent about groups. One is that there can be any number of objects in a group. We have a special name for the number of objects in the group's set. The *Order* of a group is the number of elements in it.

The order of  $(\mathbb{Z}, +)$  is infinite (there are an infinite number of integers), as is the order of  $(\mathbb{R}, +)$  and  $(\mathbb{R} - \{0\}, \cdot)$ . But, the order of  $(\{1\}, \cdot)$  is 1, and the order of  $(\mathbb{Z}_3, +)$  is 3.

If the order of a group is finite, the group is said to be *Finite*. Otherwise it is *Infinite*.

It is also clear that the elements of groups may be *Discrete*, or they may be *Continuous*. For example,  $(\mathbb{Z}, +)$ ,  $(\{1\}, \cdot)$ , and  $(\mathbb{Z}_3, +)$  are all discrete, while  $(\mathbb{R}, +)$  and  $(\mathbb{R} - \{0\}, \cdot)$  are both continuous.

Now that we understand what a discrete finite group is, we can talk about how to organize one. Namely, we use what is called a *Multiplication Table*.

A multiplication table is a way of organizing the elements of a group as follows:

$(G, \star)$	$e$	$g_1$	$g_2$	$\cdots$
$e$	$e \star e$	$e \star g_1$	$e \star g_2$	$\cdots$
$g_1$	$g_1 \star e$	$g_1 \star g_1$	$g_1 \star g_2$	$\cdots$
$g_2$	$g_2 \star e$	$g_2 \star g_1$	$g_2 \star g_2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We state the following property of multiplication tables without proof. *A multiplication table must contain every element of the group exactly one time in every row and every column.* A few minutes thought should convince you that this is necessary to ensure that the definition of a group is satisfied.

As an example, we will draw a multiplication table for the group of order 2. We won't look at specific numbers, but rather call the elements  $g_1$  and  $g_2$ . We begin as follows:

$(G, \star)$	$e$	$g_1$
$e$	?	?
$g_1$	?	?

Three of these are easy to fill in from the identity:

$(G, \star)$	$e$	$g_1$
$e$	$e$	$g_1$
$g_1$	$g_1$	?

And because we know that every element must appear exactly once, the final question mark must be  $e$ . So, there is only one possible group of order 2.

We will consider a few more examples, but we stress at this point that the temptation to plug in numbers should be avoided. Groups are abstract things, and you should try to think of them in terms of the abstract properties, not in terms of actual numbers.

We can proceed with the multiplication table for the group of order 3. You will find that, once again, there is only one option. (Doing this is instructive and it would be helpful to work this out yourself.)

$(G, \star)$	$e$	$g_1$	$g_2$
$e$	$e$	$g_1$	$g_2$
$g_1$	$g_1$	$g_2$	$e$
$g_2$	$g_2$	$e$	$g_1$

## 13.2 Group Actions

So far we have only considered elements of groups and how they relate to each other. The point has been that a particular group represents nothing more than a structure. There are a set of things, and they relate to each other in a particular way. Now, however, we want to consider an extremely simple version of how this relates to nature.

### Example 6

Consider three Easter eggs, all painted different colors (red, orange, and yellow), which we denote R, O, and Y. Now, assume they have been put into a row in the order (ROY). If we want to keep them lined up, not take any eggs away, and not add any eggs, what we can we do to them?

We can do any of the following:

1. Let  $e$  be doing nothing to the set, so  $e(ROY) = (ROY)$ ;
2. Let  $g_1$  be a cyclic permutation of the three,  $g_1(ROY) = (OYR)$ ;
3. Let  $g_2$  be a cyclic permutation in the other direction,  $g_2(ROY) = (YRO)$ ;
4. Let  $g_3$  be swapping the first and second,  $g_3(ROY) = (ORY)$ ;
5. Let  $g_4$  be swapping the first and third,  $g_4(ROY) = (YOR)$ ;
6. Let  $g_5$  be swapping the second and third,  $g_5(ROY) = (RYO)$ .

You will find that these 6 elements are closed, there is an identity, and each has an inverse. We should be very careful to draw a distinction between the *elements* of the group and the *objects* the group acts on. The objects in this example are the eggs, and the permutations are the results of the group action. Neither the eggs nor the permutations of the eggs are the elements of the group. The elements of the group are abstract objects which we are assigning to some operation on the eggs, resulting in a new permutation. We can draw a multiplication table:

$(G, \star)$	$e$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$
$e$	$e$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$
$g_1$	$g_1$	$g_2$	$e$	$g_5$	$g_3$	$g_4$
$g_2$	$g_2$	$e$	$g_1$	$g_4$	$g_5$	$g_3$
$g_3$	$g_3$	$g_4$	$g_5$	$e$	$g_1$	$g_2$
$g_4$	$g_4$	$g_5$	$g_3$	$g_2$	$e$	$g_1$
$g_5$	$g_5$	$g_3$	$g_4$	$g_1$	$g_2$	$e$

There is something interesting about this group. Notice that  $g_3 \star g_1 = g_4$ , whereas  $g_1 \star g_3 = g_5$ . So, we have the surprising result that in this group it is not necessarily true that  $g_i \star g_j = g_j \star g_i$ .

This leads to a new way of classifying groups. We say a group is *Abelian* if

$$g_i \star g_j = g_j \star g_i . \quad (\forall g_i, g_j \in G) \quad (13.1)$$

If a group is not Abelian, it is *Non-Abelian*.

Another term commonly used is *Commutate*. If

$$g_i \star g_j = g_j \star g_i , \quad (13.2)$$

then we say that  $g_i$  and  $g_j$  commute. So, an Abelian group is *Commutative*, whereas a Non-Abelian group is *Non-Commutative*.

The Easter egg group of order 6 above is an example of a very important type of group. It is denoted  $S_3$ , and is called the *Symmetric Group*. It is the group that takes three objects to all permutations of those three objects.

The more general group of this type is  $S_n$ , the group that takes  $n$  objects to all permutations of those objects. You can convince yourself that  $S_n$  will always have order  $n!$  ( $n$  factorial).

The idea above with the 3 eggs is that  $S_3$  is the group, while the eggs are the objects that the group acts on. The particular way an element of  $S_3$  changes the eggs around is called the *Group Action* of that element. And each element of  $S_3$  will move the eggs around while leaving them lined up. This ties in to our overarching concept of '*what doesn't change when something else changes*'. The fact that there are 3 eggs with 3 particular colors lined up doesn't change. The order they appear in, does.

### 13.3 Representations

We suggested above that you think of groups as purely abstract things rather than trying to plug in actual numbers. Now, however, we want to talk about how to see groups, or the elements of groups, in terms of specific numbers. But, we will do this in a very systematic way. The name for a specific set of numbers or objects that form a group is a *Representation*. The remainder of this section (and the next) will primarily be about group representations.

We already discussed a few simple representations when we discussed  $(\mathbb{Z}, +)$ ,  $(\mathbb{R} - \{0\}, \cdot)$ , and  $(\mathbb{Z}_3, +)$ . Let's focus on  $(\mathbb{Z}_3, +)$  for a moment (the integers mod 3, where  $e = 0$ ,  $g_1 = 1$ ,  $g_2 = 2$ , with addition). Notice that we could alternatively define  $e = 1$ ,  $g_1 = e^{2\pi i/3}$ , and  $g_2 = e^{4\pi i/3}$ , and let  $\star$  be multiplication. So, in the 'representation' with  $(0, 1, 2)$  and addition, we had for example

$$g_1 \star g_2 = (1 + 2) \bmod 3 = 3 \bmod 3 = 0 = e, \quad (13.3)$$

whereas now with the multiplicative representation we have

$$g_1 \star g_2 = e^{2\pi i/3} \cdot e^{4\pi i/3} = e^{2\pi i} = e^0 = 1 = e. \quad (13.4)$$

So the structure of the group is preserved in both representations.

We have two completely different representations of the same group. This idea of different ways of expressing the same group is of extreme importance, and we will be using it throughout the remainder of these lectures.

We now see a more rigorous way of coming up with representations of a particular group. We begin by introducing some notation. For a group  $(G, \star)$  with elements  $g_1, g_2, \dots$ , we call the Representation of that group  $D(G)$ , so that the elements of  $G$  are  $D(e), D(g_1), D(g_2)$  (where each  $D(g_i)$  is a matrix

of some dimension). We then choose  $\star$  to be matrix multiplication. So,  $D(g_i) \cdot D(g_j) = D(g_i \star g_j)$ .

It may not seem that we have done anything profound at this point, but we most definitely have. Remember above that we encouraged seeing groups as abstract things, rather than in terms of specific numbers. This is because a group is fundamentally an abstract object. A group is not a specific set of numbers, but rather a set of abstract objects with a well-defined structure telling you how those elements relate to each other.

The beauty of a representation  $D$  is that, via normal matrix multiplication, we have a sort of 'lens', made of familiar things (like numbers, matrices, or Easter eggs), through which we can see into this abstract world. And because  $D(g_i) \cdot D(g_j) = D(g_i \star g_j)$ , we aren't losing any of the structure of the abstract group by using a representation.

So now that we have some notation, we can develop a formalism to figure out exactly what  $D$  is for an arbitrary group.

We will use Dirac vector notation, where the column vector

$$\bar{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{pmatrix} = |v\rangle \quad (13.5)$$

and the row vector

$$\bar{v}^T = (v_1 \ v_2 \ v_3 \ \cdots) = \langle v| . \quad (13.6)$$

So, the dot product between two vectors is

$$\bar{v}^T \cdot \bar{u} = (v_1 \ v_2 \ v_3 \ \cdots) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \end{pmatrix} = v_1 u_1 + v_2 u_2 + v_3 u_3 + \cdots \equiv \langle v|u \rangle . \quad (13.7)$$

Now, we proceed by relating each element of a finite discrete group to one of the standard orthonormal unit vectors:

$$e \rightarrow |e\rangle = |\hat{e}_1\rangle, \quad g_1 \rightarrow |g_1\rangle = |\hat{e}_2\rangle, \quad g_2 \rightarrow |g_2\rangle = |\hat{e}_3\rangle . \quad (13.8)$$

And we define the way an element in a representation  $D(G)$  acts on these vectors to be

$$D(g_i)|g_j\rangle = |g_i \star g_j\rangle . \quad (13.9)$$

Now, we can build our representation. We will (from now on unless otherwise stated) represent the elements of a group  $G$  using matrices of various sizes, and the group operation  $\star$  will be standard matrix multiplication. The specific matrices that represent a given element  $g_k$  of our group will be given by

$$[D(g_k)]_{ij} = \langle g_i | D(g_k) | g_j \rangle . \quad (13.10)$$

As an example, consider again the group of order 2 (we wrote out the multiplication table above). First, we find the matrix representation of the identity,  $[D(e)]_{ij}$ ,

$$\begin{aligned} [D(e)]_{11} &= \langle e | D(e) | e \rangle = \langle e | e \star e \rangle = \langle e | e \rangle = 1 , \\ [D(e)]_{12} &= \langle e | D(e) | g_1 \rangle = \langle e | e \star g_1 \rangle = \langle e | g_1 \rangle = 0 , \\ [D(e)]_{21} &= \langle g_1 | D(e) | e \rangle = \langle g_1 | e \star e \rangle = \langle g_1 | e \rangle = 0 , \\ [D(e)]_{22} &= \langle g_1 | D(e) | g_1 \rangle = \langle g_1 | e \star g_1 \rangle = \langle g_1 | g_1 \rangle = 1 . \end{aligned} \quad (13.11)$$

So, the matrix representation of the identity is

$$D(e) \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \quad (13.12)$$

It shouldn't be surprising that the identity element is represented by the identity matrix.

Next we find the representation of  $D(g_1)$ :

$$\begin{aligned} [D(g_1)]_{11} &= \langle e | D(g_1) | e \rangle = \langle e | g_1 \star e \rangle = \langle e | g_1 \rangle = 0 , \\ [D(g_1)]_{12} &= \langle e | D(g_1) | g_1 \rangle = \langle e | g_1 \star g_1 \rangle = \langle e | e \rangle = 1 , \\ [D(g_1)]_{21} &= \langle g_1 | D(g_1) | e \rangle = \langle g_1 | g_1 \star e \rangle = \langle g_1 | g_1 \rangle = 1 , \\ [D(g_1)]_{22} &= \langle g_1 | D(g_1) | g_1 \rangle = \langle g_1 | g_1 \star g_1 \rangle = \langle g_1 | e \rangle = 0 . \end{aligned} \quad (13.13)$$

So, the matrix representation of  $g_1$  is

$$D(g_1) \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (13.14)$$

It is straightforward to check that this is a true representation,

$$\begin{aligned}
 e \star e &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e && \checkmark \\
 e \star g_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = g_1 && \checkmark \\
 g_1 \star e &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = g_1 && \checkmark \\
 g_1 \star g_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e && \checkmark
 \end{aligned} \tag{13.15}$$

Instead of considering the next obvious example, the group of order 3, consider the group  $S_3$  from above (the multiplication table is done above). The identity representation  $D(e)$  is easy – it is just the  $6 \times 6$  identity matrix. We encourage you to work out the representation of  $D(g_1)$  on your own, and check to see that it is

$$D(g_1) \doteq \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{13.16}$$

All 6 matrices can be found this way, and multiplying them out will confirm that they do indeed satisfy the group structure of  $S_3$ .

### 13.4 Reducibility and Irreducibility

You have probably noticed that equation (13.10) will always produce a set of  $n \times n$  matrices, where  $n$  is the order of the group. There is actually a name for this particular representation. The  $n \times n$  matrix representation of a group of order  $n$  is called the *Regular Representation*. More generally, the  $m \times m$  matrix representation of a group (of any order) is called the  *$m$ -dimensional representation*.

But, as we have seen, there is more than one representation for a given group (in fact, there are an infinite number of representations).



One thing we can immediately see is that any group that is Non-Abelian cannot have a  $1 \times 1$  matrix representation. This is because scalars ( $1 \times 1$  matrices) always commute, whereas matrices in general do not.

We saw above in equation (13.16) that we can represent the group  $S_n$  by  $n! \times n!$  matrices. Or, more generally, we can represent any group using  $m \times m$  matrices, where  $m$  equals order ( $G$ ). This is the regular representation. But it turns out that it is usually possible to find representations that are 'smaller' than the regular representation.

To pursue how this might be done, note that we are working with matrix representations of groups. In other words, we are representing groups in *linear spaces*. We will therefore be using a great deal of linear algebra to find smaller representations. This process, of finding a smaller representation, is called *Reducing* a representation.

Given an arbitrary representation of some group, the first question that must be asked is 'is there a smaller representation?' If the answer is yes, then the representation is said to be *Reducible*. If the answer is no, then it is *Irreducible*.

Before we dive into the more rigorous approach to reducibility and irreducibility, let's consider a more intuitive example, using  $S_3$ . In fact, we'll stick with our three painted Easter eggs,  $R$ ,  $O$ , and  $Y$ :

1.  $e(ROY) = (ROY)$ ;
2.  $g_1(ROY) = (OYR)$ ;
3.  $g_2(ROY) = (YRO)$ ;
4.  $g_3(ROY) = (ORY)$ ;
5.  $g_4(ROY) = (YOR)$ ;
6.  $g_5(ROY) = (RYO)$ .

We will represent the set of eggs by a column vector

$$|E\rangle = \begin{pmatrix} R \\ O \\ Y \end{pmatrix}. \quad (13.17)$$

Now, by inspection, what matrix would do to  $|E\rangle$  what  $g_1$  does to  $(ROY)$ ? In other words, how can we fill in the ?'s in

$$\begin{pmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \begin{pmatrix} R \\ O \\ Y \end{pmatrix} = \begin{pmatrix} O \\ Y \\ R \end{pmatrix} \quad (13.18)$$

to make the equality hold? A few moments thought will show that the appropriate matrix is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} R \\ O \\ Y \end{pmatrix} = \begin{pmatrix} O \\ Y \\ R \end{pmatrix}. \quad (13.19)$$

Continuing this reasoning, we can see that the rest of the matrices are

$$\begin{aligned} D(e) &\doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & D(g_1) &\doteq \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \\ D(g_2) &\doteq \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, & D(g_3) &\doteq \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ D(g_4) &\doteq \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & D(g_5) &\doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \end{aligned} \quad (13.20)$$

You can do the matrix multiplication to convince yourself that this is in fact a representation of  $S_3$ .

So, in equation (13.16), we had a  $6 \times 6$  matrix representation. Here, we have a new representation of consisting of  $3 \times 3$  matrices. We have therefore 'reduced' the representation. In the next section, we will look at more mathematically rigorous ways of reducing representations.

## 13.5 Algebraic Definitions

Before moving on, we must spend this section learning the definitions of several terms which are used in group theory.

If  $H$  is a subset of  $G$ , denoted  $H \subset G$ , such that the elements of  $H$  form a group, then we say that  $H$  forms a *Subgroup* of  $G$ . We make this more precise with examples.

**Example 7**

Consider (as usual) the group  $S_3$ , with the elements labeled as before:

1.  $g_0(ROY) = (ROY)$ ;
2.  $g_1(ROY) = (OYR)$ ;
3.  $g_2(ROY) = (YRO)$ ;
4.  $g_3(ROY) = (ORY)$ ;
5.  $g_4(ROY) = (YOR)$ ;
6.  $g_5(ROY) = (RYO)$ .

(where we are relabeling  $g_0 \equiv e$  for later convenience). The multiplication table is given above.

Notice that  $\{g_0, g_1, g_2\}$  form a subgroup. You can see this by noticing that the upper left 9 boxes in the multiplication table (the  $g_0, g_1, g_2$  rows and columns) all have only  $g_0$ 's,  $g_1$ 's, and  $g_2$ 's. So, here is a group of order 3 contained in  $S_3$ .

**Example 8**

Consider the subset of  $S_3$  consisting of  $g_0$  and  $g_3$  only. Both  $g_0$  and  $g_3$  are their own inverses, so the identity exists, and the group is closed. Therefore, we can say that  $\{g_0, g_3\} \subset S_3$  is a subgroup of  $S_3$ .

In fact, if you write out the multiplication table for  $g_0$  and  $g_3$  only, you will see that it is exactly equivalent to the group of order 2 considered above.

This means that we can say that  $S_3$  *contains* the group of order 2 (and we know from last time that there is only one such group, though there are an infinite number of representations of it). The way we understand this is that the abstract entity  $S_3$ , of which there is only one, contains the group of order 2, of which there is only one. However, the representations of  $S_3$ , of which there are an infinite number, will each contain the group of order 2 (of which there are also an infinite number of representations).

**Example 9**

Notice that the sets  $\{g_0, g_3\}$ ,  $\{g_0, g_4\}$ , and  $\{g_0, g_5\}$  (all  $\subset S_3$ ), are all the same as the group of order 2. This means that  $S_3$  actually contains exactly three copies of the group of order 2 in addition to the single copy of the group of order 3.

Again, this is speaking in terms of the abstract entity  $S_3$ . We can see this through the 'lens' of representation by the fact that any representation of  $S_3$  will contain three different copies of the group of order 2.

**Example 10**

As a final example of subgroups, there are two subgroups of any group, no matter what the group.

One is the subgroup consisting of only the identity,  $\{g_0\} \subset G$ . All groups contain this, but it is never very interesting.

Secondly,  $\forall G, G \subset G$ , and therefore  $G$  is always a subgroup of itself. We call these subgroups the 'trivial' subgroups.

We now introduce another important definition. If  $G$  is a group, and  $H$  is a subgroup of  $G$  ( $H \subset G$ ), then

- The set  $gH = \{g \star h | h \in H\}$  is called the *Left Coset* of  $H$  in  $G$ ;
- The set  $Hg = \{h \star g | h \in H\}$  is called the *Right Coset* of  $H$  in  $G$ .

There is a right (or left) coset for each element  $g \in G$ , though they are not necessarily all unique. This definition should be understood as follows; a coset is a *set* consisting of the elements of  $H$  all multiplied on the right (or left) by some element of  $G$ .

**Example 11**

For the subgroup  $H = \{g_0, g_1\} \subset S_3$  discussed above, the left cosets are

$$\begin{aligned} g_0\{g_0, g_1\} &= \{g_0 \star g_0, g_0 \star g_1\} = \{g_0, g_1\}; \\ g_1\{g_0, g_1\} &= \{g_1 \star g_0, g_1 \star g_1\} = \{g_1, g_2\}; \end{aligned}$$

$$\begin{aligned}
g_2\{g_0, g_1\} &= \{g_2 \star g_0, g_2 \star g_1\} = \{g_2, g_0\}; \\
g_3\{g_0, g_1\} &= \{g_3 \star g_0, g_3 \star g_1\} = \{g_3, g_4\}; \\
g_4\{g_0, g_1\} &= \{g_4 \star g_0, g_4 \star g_1\} = \{g_4, g_5\}; \\
g_5\{g_0, g_1\} &= \{g_5 \star g_0, g_5 \star g_1\} = \{g_5, g_3\}.
\end{aligned}$$

So, the left cosets of  $\{g_0, g_1\}$  in  $S_3$  are  $\{g_0, g_1\}$ ,  $\{g_1, g_2\}$ ,  $\{g_2, g_0\}$ ,  $\{g_3, g_4\}$ ,  $\{g_4, g_5\}$ , and  $\{g_5, g_3\}$ .

We can now understand the following definition.  $H$  is a *Normal Subgroup* of  $G$  if  $\forall h \in H, g^{-1} \star h \star g \in H$ . Or, in other words, if  $H$  denotes the subgroup, it is a normal subgroup if  $gH = Hg$ , which says that the left and right cosets are all equal.

As a comment, saying  $gH$  and  $Hg$  are equal doesn't mean that each individual element in the coset  $gH$  is equal to the corresponding element in  $Hg$ , but rather that the two cosets contain the same elements, regardless of their order. For example, if we had the cosets  $\{g_i, g_j, g_k\}$  and  $\{g_j, g_k, g_i\}$ , they would be equal because they contain the same three elements.

This definition means that if you take a subgroup  $H$  of a group  $G$ , and you multiply the entire set on the left by some element of  $g \in G$ , the resulting set will contain the exact same elements it would if you had multiplied *on the right* by the same element  $g \in G$ . Here is an example to illustrate.

### Example 12

Consider the order 2 subgroup  $\{g_0, g_3\} \subset S_3$ . Multiplying on the left by, say,  $g_4$ , gives

$$g_4 \star \{g_0, g_3\} = \{g_4 \star g_0, g_4 \star g_3\} = \{g_4, g_2\}. \quad (13.21)$$

And multiplying on the right by  $g_4$  gives

$$\{g_0, g_3\} \star g_4 = \{g_0 \star g_4, g_3 \star g_4\} = \{g_4, g_1\}. \quad (13.22)$$

So, because the final sets do not contain the same elements,  $\{g_4, g_2\} \neq \{g_4, g_1\}$ , we conclude that the subgroup  $\{g_0, g_3\}$  is not a normal subgroup of  $S_3$ .

### Example 13

Above, we found that  $\{g_0, g_1, g_2\} \subset S_3$  is a subgroup of order 3 in  $S_3$ . To use a familiar label, remember that we previously called the group of order 3  $(\mathbb{Z}_3, +)$ . So, dropping the '+' we refer to the group of order 3 as  $\mathbb{Z}_3$ .

#### Example 14

Consider the group of integers under addition,  $(\mathbb{Z}, +)$ . And, consider the subgroup  $\mathbb{Z}_{\text{even}} \subset \mathbb{Z}$ , the even integers under addition. Now, take some odd integer  $n_{\text{odd}}$  and act on the left:

$$n_{\text{odd}} + \mathbb{Z}_{\text{even}} = \{n_{\text{odd}} + 0, n_{\text{odd}} \pm 2, n_{\text{odd}} \pm 4, \dots\} \quad (13.23)$$

and then on the right:

$$\mathbb{Z}_{\text{even}} + n_{\text{odd}} = \{0 + n_{\text{odd}}, \pm 2 + n_{\text{odd}}, \pm 4 + n_{\text{odd}}, \dots\} . \quad (13.24)$$

Notice that the final sets are the same (because addition is commutative). So,  $\mathbb{Z}_{\text{even}} \subset \mathbb{Z}$  is a normal subgroup.

With a little thought, you can convince yourself that all subgroups of Abelian groups are normal.

If  $G$  is a group and  $H \subset G$  is normal, then the *Factor Group* of  $H$  in  $G$ , denoted  $G/H$  (read 'G mod H'), is the group with elements in the set  $G/H \equiv \{gH | g \in G\}$ . The group operation  $\star$  is understood to be

$$(g_i H) \star (g_j H) = (g_i \star g_j) H . \quad (13.25)$$

#### Example 15

Consider again  $\mathbb{Z}_{\text{even}}$ . Notice that we can call  $\mathbb{Z}_{\text{even}} = 2\mathbb{Z}$  because  $2\mathbb{Z} = 2\{0, \pm 1, \pm 2 \pm 3, \dots\} = \{0, \pm 2, \pm 4, \dots\} = \mathbb{Z}_{\text{even}}$ . We know that  $2\mathbb{Z} \subset \mathbb{Z}$  is normal, so we can build the factor group  $\mathbb{Z}/2\mathbb{Z}$  as

$$\mathbb{Z}/2\mathbb{Z} = \{0 + 2\mathbb{Z}, \pm 1 + 2\mathbb{Z}, \pm 2 + 2\mathbb{Z}, \dots\} . \quad (13.26)$$

But, notice that

$$\begin{aligned} n_{\text{even}} + 2\mathbb{Z} &= \mathbb{Z}_{\text{even}} , \\ n_{\text{odd}} + 2\mathbb{Z} &= \mathbb{Z}_{\text{odd}} . \end{aligned} \quad (13.27)$$

So, the group  $\mathbb{Z}/2\mathbb{Z}$  only has 2 elements; the set of all even integers, and the set of all odd integers. And we know from before that there is only one group of order 2, which we denote  $\mathbb{Z}_2$ . So, we have found that  $\mathbb{Z}/2\mathbb{Z} = \mathbb{Z}_2$ .

### Example 16

Finally, we consider the factor groups  $G/G$  and  $G/e$ .

- $G/G$  – The set  $G = \{g_0, g_1, g_2, \dots\}$  will be the same coset for any element of  $G$  multiplied by it. Therefore this factor group consists of only one element, and therefore  $G/G = e$ , the trivial group.
- $G/e$  — The set  $\{e\}$  will be a unique coset for any element of  $G$ , and therefore  $G/e = G$ .

Something that might help you understand factor groups better is this: the factor group  $G/H$  is the group that is left over when everything in  $H$  is 'collapsed' to the identity element. If  $G$  and  $H$  are both groups (not necessarily related in any way), then we can form the *Product Group*, denoted  $K \equiv G \otimes H$ , where an arbitrary element of  $K$  is  $(g_i, h_j)$ . If the group operation of  $G$  is  $\star_G$ , and the group operation of  $H$  is  $\star_H$ , then two elements of  $K$  are multiplied according to the rule

$$(g_i, h_j) \star_K (g_k, h_l) \equiv (g_i \star_G g_k, h_j \star_H h_l) . \quad (13.28)$$

## 13.6 Reducibility Revisited

Now that we understand subgroups, cosets, normal subgroups, and factor groups, we can begin a more formal discussion of reducing representations. Recall that in deriving equation (13.10), we made the designation

$$g_0 \rightarrow |\hat{e}_1\rangle , \quad g_1 \rightarrow |\hat{e}_2\rangle , \quad g_2 \rightarrow |\hat{e}_3\rangle , \quad \dots . \quad (13.29)$$

This was used to create an  $\text{order}(G)$ -dimensional Euclidian space which, while obviously not possessing any structure similar to the group, was and will continue to be of great use to us.

We have an  $n$ -dimensional space spanned by the orthonormal vectors  $|g_0\rangle, |g_1\rangle, \dots, |g_{n-1}\rangle$ , where  $g_0$  is understood to always refer to the identity element. This brings us to the first definition of this section. For a group  $G = \{g_0, g_1, g_2, \dots\}$ , we call the *Algebra* of  $G$  the set

$$\mathbb{C}[\mathbf{G}] \equiv \left\{ \sum_{i=0}^{n-1} c_i |g_i\rangle \mid c_i \in \mathbb{C} \forall i \right\}. \quad (13.30)$$

In other words,  $\mathbb{C}[\mathbf{G}]$  is the set of all possible linear combinations of the vectors  $|g_i\rangle$  with complex coefficients.

We could have defined the algebra over  $\mathbb{Z}$  or  $\mathbb{R}$ , but we used  $\mathbb{C}$  for generality at this point.

Addition of two elements of  $\mathbb{C}[\mathbf{G}]$  is merely normal addition of linear combinations,

$$\sum_{i=0}^{n-1} c_i |g_i\rangle + \sum_{i=0}^{n-1} d_i |g_i\rangle = \sum_{i=0}^{n-1} (c_i + d_i) |g_i\rangle. \quad (13.31)$$

This definition amounts to saying that, in the  $n$ -dimensional Euclidian space we have created, with  $n = \text{order}(G)$ , you can choose any point in the space with complex coefficients, and this will correspond to a particular linear combination of elements of  $G$ .

Now that we have defined an algebra, we can talk about group actions. Recall that the  $g_i$ 's don't act on the  $|g_j\rangle$ 's, but rather the representation  $D(g_i)$  does. We define the action  $D(g_i)$  on an element of  $\mathbb{C}[\mathbf{G}]$  as follows:

$$\begin{aligned} D(g_i) \cdot \sum_{j=0}^{n-1} c_j |g_j\rangle &= D(g_i) \cdot (c_0 |g_0\rangle + c_1 |g_1\rangle + \dots + c_{n-1} |g_{n-1}\rangle) = \\ &= c_0 |g_i \star g_0\rangle + c_1 |g_i \star g_1\rangle + \dots + c_{n-1} |g_i \star g_{n-1}\rangle = \sum_{j=0}^{n-1} c_j |g_i \star g_j\rangle. \end{aligned}$$

Previously, we discussed how elements of a group act on each other, and we also talked about how elements of a group act on some other object or set of objects (like three painted eggs). We now generalize this notion to a set of  $q$  abstract objects a group can act on, denoted  $M = \{m_0, m_1, m_2, \dots, m_{q-1}\}$ . Just as before, we build a vector space, similar to the one above used in



building an algebra. The orthonormal vectors here will be

$$m_0 \rightarrow |m_0\rangle, \quad m_1 \rightarrow |m_1\rangle, \quad \dots \quad m_{q-1} \rightarrow |m_{q-1}\rangle, \quad (13.32)$$

This allows us to understand the following definition. The set

$$\mathbf{CM} \equiv \left\{ \sum_{i=0}^{q-1} c_i |m_i\rangle \mid c_i \in \mathbb{C} \forall i \right\} \quad (13.33)$$

is called the *Module* of  $M$ . (We don't use the square brackets here to distinguish modules from algebras). In other words, the space spanned by the  $|m_i\rangle$  is the module.

### Example 17

Consider, once again,  $S_3$ . However, we generalize from three eggs to three 'objects'  $m_0$ ,  $m_1$ , and  $m_2$ . So,  $\mathbf{CM}$  is all points in the 3-dimensional space of the form  $c_0|m_0\rangle + c_1|m_1\rangle + c_2|m_2\rangle$  with  $c_i \in \mathbb{C} \forall i$ .

Then, operating on a given point with, say,  $g_1$  gives

$$\begin{aligned} g_1(c_0|m_0\rangle + c_1|m_1\rangle + c_2|m_2\rangle) &= \\ = (c_0|g_1m_0\rangle + c_1|g_1m_1\rangle + c_2|g_1m_2\rangle) \end{aligned}$$

and from the multiplication table we know

$$g_1m_0 = m_1, \quad g_1m_1 = m_0, \quad g_1m_2 = m_2.$$

So,

$$\begin{aligned} (c_0|g_1m_0\rangle + c_1|g_1m_1\rangle + c_2|g_1m_2\rangle) &= \\ = c_1|m_0\rangle + c_0|m_1\rangle + c_2|m_2\rangle. \end{aligned}$$

So, the effect of  $g_1$  was to swap  $c_1$  and  $c_0$ . This can be visualized geometrically as a reflection in the  $c_0 = c_1$  plane in the 3-dimensional module space. We can visualize every element of  $G$  in this way. They each move points around the module space in a well-defined way.

This allows us to give the following definition. If  $\mathbf{CV}$  is a module, and  $\mathbf{CW}$  is a subspace of  $\mathbf{CV}$  that is closed under the action of  $G$ , then  $\mathbf{CW}$  is an *Invariant Subspace* of  $\mathbf{CV}$ .

**Example 18**

We know that  $S_3$  acts on a 3-dimensional space spanned by

$$|m_0\rangle = (1, 0, 0)^T, \quad |m_1\rangle = (0, 1, 0)^T, \quad |m_2\rangle = (0, 0, 1)^T.$$

Now, consider the subspace spanned by

$$c(|m_0\rangle + |m_1\rangle + |m_2\rangle), \quad (13.34)$$

where  $c \in \mathbb{C}$ , and  $c$  ranges over all possible complex numbers. If we restrict  $c$  to  $\mathbb{R}$ , we can visualize this as the set of all points in the line through the origin defined by  $\lambda(\hat{i} + \hat{j} + \hat{k})$  (where  $\lambda \in \mathbb{R}$ ). You can write out the action of any element of  $S_3$  on any point in this subspace, and you will see that they are unaffected. This means that the space spanned by (13.34) is an invariant subspace.

As a note, all modules  $\mathbb{C}V$  have two trivial invariant subspaces.

- $\mathbb{C}V$  is a trivial invariant subspace of  $\mathbb{C}V$ ;
- $\mathbb{C}e$  is a trivial invariant subspace of  $\mathbb{C}V$ .

Finally, we can give a more formal definition of reducibility. If a representation  $D$  of a group  $G$  acts on the space of a module  $\mathbb{C}M$ , then the representation  $D$  is said to be *Reducible* if  $\mathbb{C}M$  contains a non-trivial invariant subspace. If a representation is not reducible, it is *Irreducible*.

In the regular representation of  $S_3$ , you may have noticed that every  $6 \times 6$  matrix appeared with non-zero elements only in the upper left  $3 \times 3$  elements, and the lower right  $3 \times 3$  elements. The upper right and lower left are all 0. This means that, for every element of  $S_3$ , there will never be any mixing of the first 3 dimensions with the last 3. So, there are two 3-dimensional invariant subspaces in the module for this particular representation of  $S_3$ .

We can now begin to take advantage of the fact that representations live in linear spaces with the following definition.

If  $V$  is any  $n$ -dimensional space spanned by  $n$  linearly independent basis vectors, and  $U$  and  $W$  are both subspaces of  $V$ , then we say that  $V$  is the *Direct Sum* of  $U$  and  $W$  if every vector  $\bar{v} \in V$  can be written as the sum

$\bar{v} = \bar{u} + \bar{w}$ , where  $\bar{u} \in U$  and  $\bar{w} \in W$ , and every operator  $X$  acting on elements of  $V$  can be separated into parts acting individually on  $U$  and  $W$ . The notation for this is  $V = U \oplus W$ .

In order to make this clearer, if  $X_n$  is an  $n \times n$  matrix, it is the direct sum of  $m \times m$  matrix  $A_m$  and  $k \times k$  matrix  $B_k$ , denoted  $X_n = A_m \oplus B_k$ , if  $X$  is in *Block Diagonal* form,

$$X_n = \begin{pmatrix} A_m & 0 \\ 0 & B_k \end{pmatrix}, \quad (13.35)$$

where  $n = m + k$ , and  $A_m$ ,  $B_k$ , and the 0's are understood as matrices of appropriate dimension.

We can generalize the previous definition as follows,

$$X_n = A_{n_1} \oplus B_{n_2} \oplus \cdots \oplus C_{n_k} = \begin{pmatrix} A_{n_1} & 0 & \cdots & 0 \\ 0 & B_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & C_{n_k} \end{pmatrix}, \quad (13.36)$$

where  $n = n_1 + n_2 + \cdots + n_k$ .

### Example 19

Let

$$A_3 = \begin{pmatrix} 1 & 1 & -2 \\ -1 & 5 & \pi \\ -17 & 4 & 11 \end{pmatrix}, \quad (13.37)$$

and let

$$B_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}. \quad (13.38)$$

Then,

$$B_2 \oplus A_3 = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & -1 & 5 & \pi \\ 0 & 0 & -17 & 4 & 11 \end{pmatrix}. \quad (13.39)$$

To summarize, we have talked about algebras, which are the vector spaces spanned by the elements of a group, and about modules, which are the vector

spaces that representations of groups act on. We have also defined invariant subspaces: Given some space and some group that acts on that space, moving the points around in a well-defined way, an invariant subspace is a subspace which always contains the same points. The group doesn't remove or add any points to that subspace. It merely moves the points around *inside* that subspace. Then, we defined a representation as reducible if there are any non-trivial invariant subspaces in the space that the group acts on.

What this amounts to is the following: a representation of any group is reducible if it can be written in block diagonal form. But this leaves the question of what we mean when we say 'can be written'. How can you 'rewrite' a representation? This leads us to the following definition. Given a matrix  $D$  and a non-singular matrix  $S$ , the linear transformation,

$$D \rightarrow D' = S^{-1}DS, \quad (13.40)$$

is called a *Similarity Transformation*.

Then, we can give the following definition. Two matrices related by a similarity transformation are said to be *Equivalent*.

Because similarity transformations are linear transformations, if  $D(G)$  is a representation of  $G$ , then so is  $S^{-1}DS$  for literally any non-singular matrix  $S$ . To see this, if  $g_i \star g_j = g_k$ , then  $D(g_i)D(g_j) = D(g_k)$ , and therefore

$$S^{-1}D(g_i)S \cdot S^{-1}D(g_j)S = S^{-1}D(g_i)D(g_j)S = S^{-1}D(g_k)S. \quad (13.41)$$

So, if we have a representation that isn't in block diagonal form, how can we figure out if it is reducible? We must look for a matrix  $S$  that will transform it into block diagonal form.

**Exercise 13.1:** Draw multiplication table for the group of order 4 (*Hint: there are 4 possibilities*).

**Exercise 13.2:** Write out elements of the multiplication table displayed for the Example 6 in Sec. 13.2.

**Exercise 13.3:** Is the subgroup  $\mathbb{Z}_3$  of  $S_3$  normal?

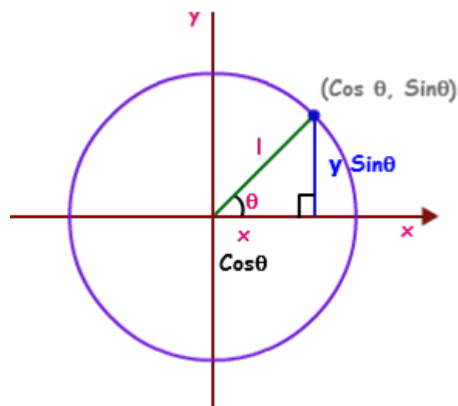
**Exercise 13.4:** Show that the subset,  $\mathbb{Z}_{\text{even}}$  (the even integers), of the group of integers under addition,  $(\mathbb{Z}, +)$ , forms the group.

**Exercise 13.5:** Show that  $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$ .

## Chapter 14

# Lie Groups

Above we considered groups which are of finite order and discrete, which allowed us to write out a multiplication table. Now we want to examine a different type of group. Consider the unit circle, where each point on the circle is specified by an angle  $\theta$ , measured from the positive  $x$ -axis.



We will refer to the point at  $\theta = 0$  as the 'starting point' (like *ROY* was for the Easter eggs). Now, just as we considered all possible orientations of (*ROY*) that left the eggs lined up, we consider all possible rotations the wheel can undergo. With the eggs there were only 6 possibilities. Now however, for the wheel there are an infinite number of possibilities for  $\theta$  (any real number  $\in [0, 2\pi)$ ).

And note that if we denote the set of all angles as  $G$ , then all the rotations obey closure ( $\theta_1 + \theta_2 = \theta_3 \in G$ ,  $\forall \theta_1, \theta_2 \in G$ ), associativity (as usual), identity ( $0 + \theta = \theta + 0 = \theta$ ), and inverse (the inverse of  $\theta$  is  $-\theta$ ).

So, we have a group that is *parameterized* by a continuous variable  $\theta$ . So, we are no longer talking about  $g_i$ 's, but about  $g(\theta)$ .

Notice that this particular group (the circle) is Abelian, which is why we can (temporarily) use addition to represent it. Also, note that we obviously cannot make a multiplication table because the order of this group is  $\infty$ .

One simple representation is the one we used above: taking  $\theta$  and using addition. A more familiar (and useful) representation is the Euler matrix,

$$g(\theta) \doteq \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (14.1)$$

with the usual matrix multiplication:

$$\begin{aligned} & \begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ -\sin \theta_1 & \cos \theta_1 \end{pmatrix} \begin{pmatrix} \cos \theta_2 & \sin \theta_2 \\ -\sin \theta_2 & \cos \theta_2 \end{pmatrix} = \\ & = \begin{pmatrix} \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 & \cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2 \\ -\sin \theta_1 \cos \theta_2 - \cos \theta_1 \sin \theta_2 & -\sin \theta_1 \sin \theta_2 + \cos \theta_1 \cos \theta_2 \end{pmatrix} = (14.2) \\ & = \begin{pmatrix} \cos(\theta_1 + \theta_2) & \sin(\theta_1 + \theta_2) \\ -\sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{pmatrix}. \end{aligned}$$

This will prove to be a much more useful representation than  $\theta$  with addition.

Groups that are parameterized by one or more continuous variables like this are called *Lie Groups*. Of course, the true definition of a Lie group is much more rigorous (and complicated), and that definition should eventually be understood. However, the definition we have given will suffice for the purposes of these notes.

## 14.1 Classification of Lie Groups

The usefulness of group theory is that groups represent a mathematical way to make changes to a system while leaving something about the system unchanged. For example, we moved (*ROY*) around, but the structure '3 eggs with different colors lined up' was preserved. With the circle, we rotated it,

but it still maintained its basic structure as a circle. It is in this sense that group theory is a study of *Symmetry*. No matter which of 'these' transformations you do to the system, 'this' stays the same – this is symmetry.

To see the usefulness of this in physics, recall Noether's Theorem (section 1.2). When you do a *symmetry* transformation to a Lagrangian, you get a conserved quantity. Recall that the transformation

$$x \rightarrow x + \epsilon \tag{14.3}$$

was a symmetry because  $\epsilon$  could take any value, and the Lagrangian was unchanged (note that  $\epsilon$  forms the Abelian group  $(\mathbb{R}, +)$ ).

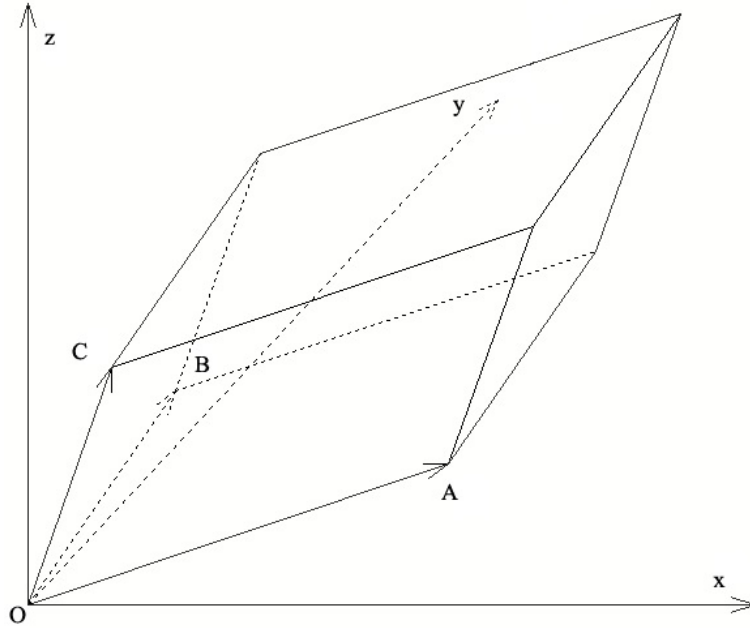
So, given a Lagrangian, which represents the structure of a physical system, a symmetry represents a way of changing the Lagrangian while preserving that structure. The particular preserved part of the system is the conserved quantity  $j$  we discussed in sections 1.2. And as you have no doubt noticed, nearly all physical processes are governed by *Conservation Laws*: conservation of momentum, energy, charge, spin, etc.

So, group theory, and in particular Lie group theory, gives us an extremely powerful way of understanding and classifying symmetries, and therefore conserved charges. And because it allows us to understand conserved charges, group theory can be used to understand the entirety of the physics in our universe.

We now begin to classify the major types of Lie groups we will be working with in these notes. To start, we consider the most general possible Lie group in an arbitrary number of dimensions,  $n$ . This will be the group that, for any point  $p$  in the  $n$ -dimensional space, can continuously take it anywhere else in the space. All that is preserved is that the points in the space stay in the space. This means that we can have literally any  $n \times n$  matrix, or *linear* transformation, as long as the matrix is invertible (non-singular). Thus, in  $n$  dimensions the largest and most general Lie group is the group of all  $n \times n$  non-singular matrices. We call this group  $GL(n)$ , or the *General Linear* group. The most general field of numbers to take the elements of  $GL(n)$  from is  $\mathbb{C}$ , so we begin with  $GL(n, \mathbb{C})$ . This is the group of all  $n \times n$  non-singular matrices with complex elements. The preserved quantity is that all points in  $\mathbb{C}^n$  stay in  $\mathbb{C}^n$ .

The most obvious subgroup of  $GL(n, \mathbb{C})$  is  $GL(n, \mathbb{R})$ , or the set of all  $n \times n$  invertible matrices with real elements. This leaves all points in  $\mathbb{R}^n$  in  $\mathbb{R}^n$ .

To find a further subgroup, recall from linear algebra and vector calculus that in  $n$  dimensions, you can take  $n$  vectors at the origin such that for a parallelepiped, we could obtain:



Then, if you arrange the components of the  $n$  vectors into the rows (or columns) of a matrix, the determinant of that matrix will be the volume of the parallelepiped.

So, consider now the set of all General Linear transformations that transform all vectors from the origin (or in other words, points in the space) in such a way that the volume of the corresponding parallelepiped is preserved. This will demand that we only consider General Linear matrices with determinant equal to 1. Also, the set of all General Linear matrices with unit determinant will form a group because of the general rule

$$\det |A \cdot B| = \det |A| \cdot \det |B| . \quad (14.4)$$

So, if  $\det |A| = 1$  and  $\det |B| = 1$ , then  $\det |A \cdot B| = 1$ . We call this subgroup of  $GL(n, \mathbb{C})$  the *Special Linear* group, or  $SL(n, \mathbb{C})$ . The natural subgroup of this is  $SL(n, \mathbb{R})$ . This group preserves not only the points in the space (as  $GL$  did), but also the volume, as described above.



Now, consider the familiar transformations on vectors in  $n$ -dimensional space of generalized Euler angles. These are transformations that rotate all points around the origin. These rotation transformations leave the radius squared ( $r^2$ ) invariant. And, because  $\bar{r}^2 = \bar{r}^T \cdot \bar{r}$ , if we transform with a rotation matrix  $R$ , then  $\bar{r} \rightarrow \bar{r}' = R\bar{r}$ , and  $\bar{r}^T \rightarrow \bar{r}'^T = \bar{r}^T R^T$ , so  $\bar{r}'^T \cdot \bar{r}' = \bar{r}^T R^T \cdot R\bar{r}$ . But, as we said, we are demanding that the radius squared be invariant under the action of  $R$ , and so we demand  $\bar{r}^T R^T \cdot R\bar{r} = \bar{r}^T \cdot \bar{r}$ . So, the constraint we are imposing is  $R^T \cdot R = \mathbb{I}$ , which implies  $R^T = R^{-1}$ . This tells us that the rows and columns of  $R$  are orthogonal. Therefore, we call the group of generalized rotations, or generalized Euler angles in  $n$  dimensions,  $O(n)$ , or the *Orthogonal* group. We don't specify  $\mathbb{C}$  or  $\mathbb{R}$  here because it will be understood that we are always talking about  $\mathbb{R}$ .

Also, note that because

$$\det |R^T \cdot R| = \det |\mathbb{I}|, \quad \Rightarrow \quad (\det |R|)^2 = 1, \quad \Rightarrow \quad \det |R| = \pm 1. \quad (14.5)$$

We again denote the subgroup with  $\det |R| = +1$  the *Special Orthogonal* group, or  $SO(n)$ . To understand what this means, consider an orthogonal matrix with determinant  $-1$ , such as

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (14.6)$$

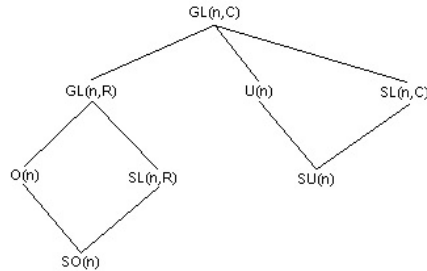
This matrix is orthogonal, and therefore is an element of the group  $O(3)$ , but the determinant is  $-1$ . This matrix will take the point  $(x, y, z)^T$  to the point  $(x, y, -z)^T$ . This changes the handedness of the system (the right hand rule will no longer work). So, if we limit ourselves to  $SO(n)$ , we are preserving the space, the radius, the volume, and the handedness of the space.

For vectors in  $\mathbb{C}$  space, we do not define orthogonal matrices (although we could). Instead, we discuss the complex version of the radius, where instead of  $\bar{r}^2 = \bar{r}^T \cdot \bar{r}$ , we have  $\bar{r}^2 = \bar{r}^\dagger \cdot \bar{r}$ , where the dagger denotes the Hermitian conjugate,  $\bar{r}^\dagger = (\bar{r}^\star)^T$ , where  $\star$  denotes complex conjugate.

So, with the elements in  $R$  being in  $\mathbb{C}$ , we have  $\bar{r} \rightarrow R\bar{r}$ , and  $\bar{r}^\dagger \rightarrow \bar{r}^\dagger R^\dagger$ . So,  $\bar{r}^\dagger \cdot \bar{r} \rightarrow \bar{r}^\dagger R^\dagger \cdot R\bar{r}$ , and by the same argument as above with the orthogonal matrices, this demands that  $R^\dagger \cdot R = \mathbb{I}$ , or  $R^\dagger = R^{-1}$ . We denote such matrices *Unitary*, and the set of all such  $n \times n$  invertible matrices form the group  $U(n)$ . Again, we understand the unitary groups to have elements in  $\mathbb{C}$ ,

so we don't specify that. And, we will still have a subset of unitary matrices  $R$  with  $\det |R| = 1$  called  $SU(n)$ , the *Special Unitary* groups.

We can summarize the hierarchy we have just described in the following diagram:



We will now describe one more category of Lie groups before moving on. We saw above that the group  $SO(n)$  preserves the radius squared in real space. In coordinates, this means that  $\bar{r}^2 = x_1^2 + x_2^2 + \cdots + x_n^2$ , or more generally the dot product  $\bar{x} \cdot \bar{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$  is preserved.

However, we can generalize this to form a group action that preserves not the radius squared, but the value (switching to initial notation for the dot product)

$$x^a y_a = -x_1y_1 - x_2y_2 - \cdots - x_my_m + x_{m+1}y_{m+1} + \cdots + x_{m+n}y_{m+n} . \quad (14.7)$$

We call the group that preserves this quantity  $SO(m, n)$ . The space we are working in is still  $\mathbb{R}^{m+n}$ , but we are making transformations that preserve something different than the radius.

Note that  $SO(m, n)$  will have an  $SO(m)$  subgroup and an  $SO(n)$  subgroup, consisting of rotations in the first  $m$  and last  $n$  components separately.

Finally, notice that the specific group of this type,  $SO(1, 3)$ , is the group that preserves the value  $s^2 = -x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$ , or written more suggestively,  $s^2 = -c^2t^2 + x^2 + y^2 + z^2$ . Therefore, the group  $SO(1, 3)$  is the *Lorentz Group*. Any action that is invariant under  $SO(1, 3)$  is said to be a Lorentz Invariant theory (as all theories should be). We will find that thinking of Special Relativity in these terms will be much more useful.

It should be noted that there are many other types of Lie groups. We have limited ourselves to the ones we will be working with in these notes.

## 14.2 Generators

Now that we have a good 'birds eye view' of Lie groups, we can begin to pick apart the details of how they work.

As we said before, a Lie group is a group that is parameterized by a set of continuous parameters, which we call  $\alpha_i$  for  $i = 1, \dots, n$ , where  $n$  is the number of parameters the group depends on. The elements of the group will then be denoted  $g(\alpha_i)$ .

Because all groups include an identity element, we will choose to parameterize them in such a way that  $g(\alpha_i)|_{\alpha_i=0} = e$ , the identity element. So, if we are going to talk about representations,  $D_n(g(\alpha_i))|_{\alpha_i=0} = \mathbb{I}$ , where  $\mathbb{I}$  is the  $n \times n$  identity matrix for whatever dimension ( $n$ ) representation we want.

Now, take  $\alpha_i$  to be very small with  $\delta\alpha_i \ll 1$ . So,  $D_n(g(0 + \delta\alpha_i))$  can be Taylor expanded:

$$D_n(g(\delta\alpha_i)) = \mathbb{I} + \delta\alpha_i \frac{\partial D_n(g(\alpha_i))}{\partial \alpha_i} \Big|_{\alpha_i=0} + \dots \quad (14.8)$$

The terms  $\partial D_n / \partial \alpha_i|_{\alpha_i=0}$  are extremely important, and we give them their own expression:

$$X_i \equiv -i \frac{\partial D_n}{\partial \alpha_i} \Big|_{\alpha_i=0} \quad (14.9)$$

(we have included the  $-i$  in order to make  $X_i$  Hermitian, which will be necessary later).

So, the representation for infinitesimal  $\delta\alpha_i$  is then

$$D_n(\delta\alpha_i) = \mathbb{I} + i\delta\alpha_i X_i + \dots \quad (14.10)$$

(where we have switched our notation from  $D_n(g(\alpha))$  to  $D_n(\alpha)$  for brevity).

The  $X_i$ 's are constant matrices which we will determine later.

Now, let's say that we want to see what the representation will look like for a finite value of  $\alpha_i$  rather than an infinitesimal value. A finite transformation will be the result of an infinite number of infinitesimal transformations. Or in other words,  $\alpha_i = N\delta\alpha_i$  as  $N \rightarrow \infty$ . So,  $\delta\alpha_i = \alpha_i/N$ , and an infinite number of infinitesimal transformations is

$$\lim_{N \rightarrow \infty} (1 + i\delta\alpha_i X_i)^N = \lim_{N \rightarrow \infty} \left(1 + i \frac{\alpha_i}{N} X_i\right)^N \quad (14.11)$$

If you expand this out for several values of  $N$ , you will see that it is exactly

$$\lim_{N \rightarrow \infty} \left(1 + i \frac{\alpha_i}{N} X_i\right)^N = e^{i\alpha_i X_i} \quad (14.12)$$

We call the  $X_i$ 's the *Generators* of the group, and there is one for each parameter required to specify a particular element of the group. For example, consider  $SO(3)$ , the group of rotations in 3 dimensions. We know from vector calculus that an element of  $SO(3)$  requires 3 angles, usually denoted  $\theta$ ,  $\phi$  and  $\psi$ . Therefore,  $SO(3)$  will require 3 generators, which will be denoted  $X_\theta$ ,  $X_\phi$ , and  $X_\psi$ . We will discuss how the generators can be found soon.

In general, there will be several (in fact, infinite) different sets of  $X_i$ 's that define a given group (just as there are an infinite number of representations of any finite group). What we will find is that up to a similarity transformation, a particular set of generators defines a particular representation of a group.

So,  $D_n(\alpha_i) = e^{i\alpha_i X_i}$  for any group (the  $i$  index in the exponent is understood to be summed over all parameters and generators). The best way to think of the parameter space for the group is as a vector space, where the generators describe the behavior near the identity, but form a basis for the entire vector space. By analogy, think of the unit vectors  $\hat{i}$ ,  $\hat{j}$ , and  $\hat{k}$  in  $\mathbb{R}^3$ . They are defined at the origin, but they can be combined with real numbers/parameters to specify any arbitrary point in  $\mathbb{R}^3$ . In the same way, the generators are the 'unit vectors' of the parameter space (which in general is a much more complicated space than Euclidian space), and the parameters (like  $\theta$ ,  $\phi$ , and  $\psi$ ) specify where in the parameter space you are in terms of the generators. That point in the parameter space will then correspond to a particular element of the group.

We call the number of generators of a group (or equivalently the number of parameters necessary to specify an element), the *Dimension* of the group. For example, the dimension of  $SO(3)$  is 3. The dimension of  $SO(2)$  (rotations in the plane) however is only 1 (only  $\theta$  is needed), so there will be only one generator.

### 14.3 Lie Algebras

An algebra is a space spanned by elements of the group with  $\mathbb{C}$  coefficients parameterizing the Euclidian space we defined. Obviously we can't define an

algebra in the same way for Lie groups, because the elements are continuous. But, as discussed in the last section, a particular element of a Lie group is defined by the values of the parameters in the parameter space spanned by the generators. We will see that the generators will form the algebras for Lie groups.

Consider two elements of the same group with generators  $X_i$ , one with parameter values  $\alpha_i$  and the other with parameter values  $\beta_i$ . The product of the 2 elements will then be  $e^{i\alpha_i X_i} e^{i\beta_j X_j}$ . Because we are assuming this is a group, we know that the product must be an element of the group (due to closure), and therefore the product must be specified by some set of parameters  $\delta_k$ , so

$$e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i\delta_k X_k} . \quad (14.13)$$

Note that the product won't necessarily simply be

$$e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i(\alpha_i X_i + \beta_j X_j)} \quad (14.14)$$

because the generators are matrices and therefore don't in general commute.

So, we want to figure out what  $\delta_i$  will be in terms of  $\alpha_i$  and  $\beta_i$ . We do this as follows.

$$\begin{aligned} i\delta_i X_k &= \ln(e^{i\delta_k X_k}) = \ln(e^{i\alpha_i X_i} e^{i\beta_j X_j}) = \\ &= \ln(1 + e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1) \equiv \ln(1 + x) , \end{aligned} \quad (14.15)$$

where we have defined

$$x \equiv e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1 . \quad (14.16)$$

We will proceed by expanding only to second order in  $\alpha_i$  and  $\beta_j$ , though the result we will obtain will hold at arbitrary order. By Taylor expanding the exponential terms,

$$\begin{aligned} e^{i\alpha_i X_i} e^{i\beta_j X_j} - 1 &= \left[ 1 + i\alpha_i X_i + \frac{1}{2}(i\alpha_i X_i)^2 + \dots \right] \times \\ &\times \left[ 1 + i\beta_j X_j + \frac{1}{2}(i\beta_j X_j)^2 + \dots \right] - 1 = \\ &= 1 + i\beta_j X_j - \frac{1}{2}(\beta_j X_j)^2 + i\alpha_i X_i - \\ &- \alpha_i X_i \beta_j X_j - \frac{1}{2}(\alpha_i X_i)^2 - 1 = \\ &= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \\ &- \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2] . \end{aligned} \quad (14.17)$$

Then, using the general Taylor expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots, \quad (14.18)$$

and again keeping terms only to second order in  $\alpha$  and  $\beta$ , we have

$$\begin{aligned} x - \frac{x^2}{2} &= \left[ i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2] \right] - \\ &- \frac{1}{2} \left[ i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2] \right]^2 = \\ &= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2] - \\ &- \frac{1}{2} [-(\alpha_i X_i + \beta_j X_j)(\alpha_i X_i + \beta_j X_j)] = \\ &= i(\alpha_i X_i + \beta_j X_j) - \alpha_i X_i \beta_j X_j - \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2] + \\ &+ \frac{1}{2} [(\alpha_i X_i)^2 + (\beta_j X_j)^2 + \alpha_i \beta_j (X_i X_j + X_j X_i)] = \quad (14.19) \\ &= i(\alpha_i X_i + \beta_j X_j) + \frac{1}{2} \alpha_i \beta_j (X_j X_i - X_i X_j) = \\ &= i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} \alpha_i \beta_j [X_i, X_j] = \\ &= i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j]. \end{aligned}$$

So finally we can see

$$i\delta_k X_k = i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j], \quad (14.20)$$

or

$$e^{i\alpha_i X_i} e^{i\beta_j X_j} = e^{i(\alpha_i X_i + \beta_j X_j) - \frac{1}{2} [\alpha_i X_i, \beta_j X_j]}. \quad (14.21)$$

Equation (14.21) is called the *Baker-Campbell-Hausdorff* formula, and it is one of the most important relations in group theory and in physics. Notice that, if the generators commute, this reduces to the normal equation for multiplying exponentials. You can think of equation (14.21) as the generalization of the normal exponential multiplication rule.

Now, it is clear that the commutator  $[X_i, X_j]$  must be proportional to some linear combination of the generators of the group (because of closure). So, it must be the case that

$$[X_i, X_j] = if_{ijk} X_k \quad (14.22)$$

for some set of constants  $f_{ijk}$ . These constants are called the *Structure Constants* of the group, and if they are completely known, the commutation relations between all the generators are known, and so the entire group can be determined in any representation you want.

The generators, under the specific commutation relations defined by the structure constants, form the *Lie Algebra* of the group, and it is this commutation structure which forms the structure of the Lie group.

## 14.4 The Adjoint Representation

We will talk about several representations for each group we discuss, but we will mention a very important one now. We mentioned before that the structure constants  $f_{ijk}$  completely determine the entire structure of the group.

We begin by using the Jacobi identity,

$$[X_i, [X_j, X_k]] + [X_j, [X_k, X_i]] + [X_k, [X_i, X_j]] = 0 \quad (14.23)$$

(if you aren't familiar with this identity, try multiplying it out. You will find that it is identically true – all the terms cancel exactly). But, from equation (14.22), we can write

$$[X_i, [X_j, X_k]] = if_{jka}[X_i, X_a] = if_{jka}f_{iab}X_b . \quad (14.24)$$

Plugging this into (14.23) we get

$$\begin{aligned} if_{jka}f_{iab}X_b + if_{kia}f_{jab}X_b + if_{ija}f_{kab}X_b &= 0 , \Rightarrow \\ \Rightarrow (f_{jka}f_{iab} + f_{kia}f_{jab} + f_{ija}f_{kab})iX_b &= 0 , \Rightarrow \\ \Rightarrow f_{jka}f_{iab} + f_{kia}f_{jab} + f_{ija}f_{kab} &= 0 . \end{aligned} \quad (14.25)$$

So, if we define the matrices

$$[T^a]_{bc} \equiv -if_{abc} , \quad (14.26)$$

then it is easy to show that (14.25) leads to

$$[T^a, T^b] = if_{abc}T^c . \quad (14.27)$$

So, the structure constants themselves form a representation of the group, as defined by (14.26). We call this representation the *Adjoint Representation*, and it will prove to be extremely important.

Notice that the indices labeling the rows and columns in (14.26) each run over the same values as the indices labeling the  $T$  matrices. This tells us that the adjoint representation is made of  $n \times n$  matrices, where  $n$  is the dimension of the group, or the number of parameters in the group. For example,  $SO(3)$  requires 3 parameters to specify an element  $(\theta, \phi, \psi)$ , so the adjoint representation of  $SO(3)$  will consist of  $3 \times 3$  matrices.  $SO(2)$  on the other hand is Abelian, and therefore all of the structure constants vanish. Therefore there is no adjoint representation of  $SO(2)$ .

## 14.5 Root Space

We saw in the previous section that we can view the physical space that a group is acting on by using the eigenvectors of the diagonal generators as a basis. These eigenvectors can be arranged in order of decreasing eigenvalue. Then, the non-diagonal generators can be used to form linear combinations that act as raising and lowering operators, which transform one eigenvector to another, changing the eigenvalue by an amount defined by the commutation relations of the generators.

We now see that this generalizes very nicely.

An arbitrary Lie group is defined in terms of its generators. As we said at the end of section 14.2, it is best to think of the generators as being analogous to the basis vectors spanning some space. Of course, the space the generators span is much more complicated than  $\mathbb{R}^n$  in general, but the generators span the space the same way. In this sense, the generators form a linear vector space. So, we must define an inner product for them. For reasons that are beyond the scope of these notes, we will choose the generators and inner product so that, for generators  $T^a$  and  $T^b$ ,

$$\langle T^a, T^b \rangle \equiv \frac{1}{\kappa} \text{Tr}(T^a T^b) = \delta^{ab}, \quad (14.28)$$

where  $\kappa$  is some normalization constant.

Also, in the set of generators of a Lie group, there will be a closed subalgebra of generators which all commute with each other, but not with generators



outside of this subalgebra. In other words, this is the set of generators which can be simultaneously diagonalized through some similarity transformation. For  $SU(2)$ , we saw that there was only one generator in this subalgebra which we chose to be  $J_j^3$  (recall that a matrix will only commute with all other matrices if it is equal to the identity matrix times a constant, whereas two diagonal matrices will always commute regardless of what their diagonal elements are).

Let's say that a particular Lie group has  $N$  generators total, or is an  $N$ -dimensional group. Then, let's say that there are  $M < N$  generators in the mutually commuting subalgebra. We call those  $M$  generators the *Cartan Subalgebra*, and the generators in it are called *Cartan Generators*. We define the number  $M$  as the *Rank* of the group.

By convention we will label the Cartan generators  $H^i$  ( $i = 1, \dots, M$ ) and the non-Cartan generators  $E^i$  ( $i = 1, \dots, N - M$ ). For example, with  $SU(2)$  we had

$$H^1 = J_j^3, \quad E^1 = J_j^1, \quad E^2 = J_j^2. \quad (14.29)$$

Before moving on, we point out that this should seem familiar. If you think back to an introductory class in quantum mechanics, recall that we always choose some set of variables that all commute with each other (usually we choose either position or momentum because  $[x, p] \neq 0$ ). Then, we expand the physical states in terms of the position or momentum eigenvectors. Here, we are doing the exact same thing, only in a much more general context.

Now, the  $H^i$ 's are simultaneously diagonalized, so we will write the physical states in terms of their eigenvalues. In an  $n$ -dimensional representation  $D_n$ , the generators are  $n \times n$  matrices, so the eigenvectors are  $n$ -dimensional. So, there will be a total of  $n$  eigenvectors, and each will have one eigenvalue with each of the  $M$  Cartan generators  $H^i$ . So, for each of these eigenvectors, which we temporarily denote  $|j\rangle$ , for  $j = 1, \dots, n$ , we have the  $M$  eigenvalues with  $M$  Cartan generators, which we call  $t_j^i$  (where  $j = 1, \dots, n$  labels the eigenvectors, and  $i = 1, \dots, M$  labels the eigenvalues), and we form what is called a *Weight Vector*

$$\bar{t}_j \equiv \begin{pmatrix} t_j^1 \\ t_j^2 \\ t_j^3 \\ \vdots \\ t_j^M \end{pmatrix}, \quad (14.30)$$

where  $j = 1, \dots, n$ . The individual components of these vectors, the  $t_j^i$ 's, are called the *Weights*.

So for a given representation  $D_n$ , we now denote the state  $|D_n; \bar{t}_j\rangle$  (instead of  $|j\rangle$ ). So, our eigenvalues will be

$$H^i |D_n; \bar{t}_j\rangle = t_j^i |D_n; \bar{t}_j\rangle. \quad (14.31)$$

As we mentioned before, the adjoint representation is a particularly important representation. If you do not remember the details of the adjoint representation, go reread the section 14.4. Here, the generators are defined by equation (14.26),  $[T^a]_{bc} \equiv -if_{abc}$ . Recall that each index runs from 1 to  $N$ , so that the generators in the adjoint representation are  $N \times N$  matrices, and the eigenvectors are  $N$ -dimensional.

Also, as a point of nomenclature, weights in the adjoint representation are called *Roots*, and the corresponding vectors (as in (14.30)) are called *Root Vectors*.

This means that there is exactly one eigenvector for each generator, and therefore one root vector for each generator. So, in equation (14.30),  $j = 1, \dots, N$ . We make this more obvious by explicitly assigning each eigenvector to a generator as follows. First, because we now have the same number of generators, eigenvectors, and root vectors, we label the generators by the root vectors  $T^{\bar{t}_j}$  instead of  $T^j$ . Also, we now refer to general eigenstates as  $|Adj; T^{\bar{t}_j}\rangle$ , where  $j = 1, \dots, N$  and  $\bar{t}_j$  is the  $M$ -dimensional root vector corresponding to  $T^{\bar{t}_j}$ . And, we also divide the states  $|Adj; T^{\bar{t}_j}\rangle$  into two groups: those corresponding to the  $M$  Cartan generators  $|Adj; H^{\bar{h}_j}\rangle$  (where  $j = 1, \dots, M$  and  $\bar{h}_j$  is the  $M$ -dimensional root vector corresponding to  $H^{\bar{h}_j}$ ), and those corresponding to the  $N - M$  non-Cartan generators  $|Adj; E^{\bar{e}_j}\rangle$  (where  $j = 1, \dots, N - M$  and  $\bar{e}_j$  is the  $M$ -dimensional root vector corresponding to  $E^{\bar{e}_j}$ ).

Don't be alarmed by the superscripts being vectors. We are using this notation for later convenience, and  $T^{\bar{t}_i}$  here means the same thing  $T^j$  did before (the  $j^{\text{th}}$  generator). This notation, which we use only for the adjoint representation, is simply taking advantage of the fact that in the adjoint representation, the total number of generators, the number of eigenvectors of the Cartan generators, the dimension of the representation, and the number of weight/root vectors is the same.

Also, with the adjoint representation states  $|Adj; T^{\bar{j}}\rangle$ , we can use equation (14.28) to define the inner product between states as

$$\langle Adj; T^{\bar{j}} | Adj; T^{\bar{k}} \rangle = \frac{1}{\kappa} \text{Tr}(T^{\bar{j}} T^{\bar{k}}) = \delta^{jk} . \quad (14.32)$$

We will make use of this equation soon.

The matrix elements of a given generator will then be given by the familiar equation

$$-if_{abc} = [T^{\bar{a}}]_{bc} \equiv \langle Adj; T^{\bar{b}} | T^{\bar{a}} | Adj; T^{\bar{c}} \rangle . \quad (14.33)$$

We want to know what an arbitrary generator  $T^{\bar{a}}$  will do to an arbitrary state  $|Adj; T^{\bar{b}}\rangle$  in the adjoint representation. So,

$$\begin{aligned} T^{\bar{a}} |Adj; T^{\bar{b}}\rangle &= \sum_c |Adj; T^{\bar{c}}\rangle \langle Adj; T^{\bar{c}} | T^{\bar{a}} | Adj; T^{\bar{b}}\rangle = \\ &= \sum_c |Adj; T^{\bar{c}}\rangle [T^{\bar{a}}]_{cb} = \\ &= \sum_c |Adj; T^{\bar{c}}\rangle (-if_{acb}) = \\ &= \sum_c if_{abc} |Adj; T^{\bar{c}}\rangle . \end{aligned} \quad (14.34)$$

And, because there is exactly one eigenvector for each generator, the state  $|Adj; T^{\bar{c}}\rangle$  corresponds to the generator  $T^c$ . And because we know that

$$if_{abc} T^{\bar{c}} = [T^{\bar{a}}, T^{\bar{b}}] , \quad (14.35)$$

(where  $c$  is understood to be summed) by definition of the structure constants, we can infer that

$$T^{\bar{a}} |Adj; T^{\bar{b}}\rangle = \sum_c if_{abc} |Adj; T^{\bar{c}}\rangle = |Adj; [T^{\bar{a}}, T^{\bar{b}}]\rangle , \quad (14.36)$$

where  $[T^{\bar{a}}, T^{\bar{b}}]$  is simply the commutator.

Let's apply this to combinations of the two types of generators we have,  $H^{\bar{h}_a}$ 's and  $E^{\bar{e}_a}$ 's. If we have a Cartan generator acting on a state corresponding to a Cartan generator, we have (from equation (14.31))

$$H^{\bar{h}_a} |Adj; H^{\bar{h}_b}\rangle = h_b^a |Adj; H^{\bar{h}_b}\rangle . \quad (14.37)$$

But from (14.36) we have

$$H^{\bar{h}_a} |Adj; H^{\bar{h}_b}\rangle = |Adj; [H^{\bar{h}_a}, H^{\bar{h}_b}]\rangle . \quad (14.38)$$

By definition, the Cartan generators commute, so  $[H^{\bar{h}_a}, H^{\bar{h}_b}] \equiv 0$ , and therefore

$$\bar{h}_b \equiv 0 . \quad (14.39)$$

So we can drop them from our notation, leaving the eigenstates corresponding to non-Cartan generators denoted  $|Adj; H^j\rangle$ .

On the other hand, if we have a Cartan generator acting on an eigenstate corresponding to a non-Cartan generator, equation (14.31) gives

$$H^a |Adj; E^{\bar{e}_b}\rangle = e_b^a |Adj; [H^a, E^{\bar{e}_b}]\rangle . \quad (14.40)$$

And equation (14.36) gives

$$H^a |Adj; E^{\bar{e}_b}\rangle = |Adj; [H^a, E^{\bar{e}_b}]\rangle . \quad (14.41)$$

Now, we don't know a priori what  $[H^a, E^{\bar{e}_b}]$  is, but comparing (14.40) and (14.41), we see

$$|Adj; e_b^a E^{\bar{e}_b}\rangle = |Adj; [H^a, E^{\bar{e}_b}]\rangle . \quad (14.42)$$

And because we know that each of these vectors corresponds directly to the generators, we have the final result

$$[H^a, E^{\bar{e}_b}] = e_b^a E^{\bar{e}_b} . \quad (14.43)$$

Let us find out what a non-Cartan generator does to a given eigenstate. Consider an arbitrary state  $|Adj; T^{\bar{t}_b}\rangle$  with  $H^c$  eigenvalue  $t_b^c$ . We can act on this with  $E^{\bar{e}_a}$  to create the new state  $E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle$ . So what will the  $H^c$  eigenvalue of this new state be? Using (14.43),

$$\begin{aligned} H^c E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle &= (H^c E^{\bar{e}_a} - E^{\bar{e}_a} H^c + E^{\bar{e}_a} H^c) |Adj; T^{\bar{t}_b}\rangle = \\ &= ([H^c, E^{\bar{e}_a}] + E^{\bar{e}_a} H^c) |Adj; T^{\bar{t}_b}\rangle = \\ &= (e_a^c E^{\bar{e}_a} + E^{\bar{e}_a} t_b^c) |Adj; T^{\bar{t}_b}\rangle = \\ &= (t_b^c + e_a^c) E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle = \\ &= (\bar{t}_b + \bar{e}_a)^c E^{\bar{e}_a} |Adj; T^{\bar{t}_b}\rangle . \end{aligned} \quad (14.44)$$

So, by acting on the one of the eigenstates with a non-Cartan generator  $E^{\bar{e}_a}$ , we have shifted the  $H^c$  eigenvalue by one of the coordinates of the root vector.

What this means is that the non-Cartan generators play a role analogous to the raising and lowering operators in  $SU(2)$  (see below), except instead of merely shifting the state 'up' and 'down', it moves the states around through some  $M$ -dimensional space.

From this, we can also see that if there is an operator that can transform from one state to another, there must be a corresponding operator that will make the opposite transformation. Therefore, for every operator  $E^{\bar{e}_a}$ , we expect to have the operator  $E^{-\bar{e}_a}$ , and corresponding eigenstate  $|Adj; E^{-\bar{e}_a}\rangle$ .

Finally, consider the state  $E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle$ . We know from (14.36) that

$$E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = |Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}]\rangle . \quad (14.45)$$

The eigenvalue of this state can be found using equation (14.44):

$$H^b E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = (-\bar{e}_a + \bar{e}_a)^b E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle \equiv 0 . \quad (14.46)$$

But according to equation (14.39), states with 0 eigenvalue are states corresponding to Cartan generators. Therefore we conclude that the state  $E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle$  is proportional to some linear combination of the Cartan states,

$$E^{\bar{e}_a}|Adj; E^{-\bar{e}_a}\rangle = \sum_b N_b |Adj; H^b\rangle , \quad (14.47)$$

where the  $N_b$ 's are the constants of proportionality. To find the constants  $N_b$ , we follow an approach similar to the one we used in deriving (15.36). Taking the inner product and using (14.36),

$$\begin{aligned} \langle Adj; H^c | E^{\bar{e}_a} | Adj; E^{-\bar{e}_a} \rangle &= \sum_b N_b \langle Adj; H^c | Adj; H^b \rangle = \\ &= \sum_b N_b \delta^{cb} = N_c , \quad \Rightarrow \quad (14.48) \\ \Rightarrow \quad \langle Adj; H^c | Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}] \rangle &= N_c . \end{aligned}$$

Then, using (14.32)

$$\begin{aligned} \langle Adj; H^c | Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}] \rangle &= \frac{1}{\kappa} \text{Tr}(H^c [E^{\bar{e}_a}, E^{-\bar{e}_a}]) = \\ &= \frac{1}{\kappa} \text{Tr}(E^{-\bar{e}_a} [H^c, E^{\bar{e}_a}]) = \\ &= \frac{1}{\kappa} e_a^c \text{Tr}(E^{-\bar{e}_a} E^{\bar{e}_a}) = \quad (14.49) \\ &= e_a^c \delta^{aa} = \\ &= e_a^c . \end{aligned}$$

So,

$$N_c = e_a^c . \quad (14.50)$$

And therefore equation (14.47) is now

$$E^{\bar{e}_a} |Adj; E^{-\bar{e}_a}\rangle = |Adj; [E^{\bar{e}_a}, E^{-\bar{e}_a}]\rangle = e_a^b |Adj; H^b\rangle , \quad (14.51)$$

where the sum over  $b$  is understood. This leads to our final result,

$$[E^{\bar{e}_a}, E^{-\bar{e}_a}] = e_a^b H^b . \quad (14.52)$$

Though we did all of this using the adjoint representation we have seen before, this structure is the same in any representation, and therefore everything we have said is valid in any  $D_n$ . We worked in the adjoint simply because that makes the results easiest to obtain. The extensive use we made of labeling the eigenvectors with the generators can only be done in the adjoint representation because only in the adjoint does the number of eigenvectors equal the number of eigenstates. However, this will not be a problem.

The important results from this section are (14.43) and (14.52), which are true in any representation. Part of what we will do later is find these structures in other representations.

## 14.6 Casimir Operators

A Casimir invariant, or Casimir operator, is a distinguished element of the center of the universal enveloping algebra of a Lie algebra (universal enveloping algebra is the most general algebra that contains all representations of the Lie algebra). The Casimir element is named after Hendrik Casimir, who identified them in his description of rigid body dynamics in 1931.

The most commonly-used Casimir operator is the quadratic invariant, which is the simplest to define. However, one may also have Casimir invariants of higher order, which correspond to homogeneous symmetric polynomials of higher order.

Suppose that  $\mathfrak{g}$  is an  $n$ -dimensional semisimple Lie algebra (a Lie algebra which is a direct sum of simple Lie algebras). Let  $B$  be a nondegenerate

bilinear form on  $\mathfrak{g}$  that is invariant under the adjoint action of  $\mathfrak{g}$  on itself, meaning that

$$B(ad_X Y, Z) + B(Y, ad_X Z) = 0 \quad (14.53)$$

for all  $X, Y, Z$  in  $\mathfrak{g}$ . Let  $\{X_i\}_{i=1}^n$  be any basis of  $\mathfrak{g}$ , and  $\{X^i\}_{i=1}^n$  be the dual basis of  $\mathfrak{g}$  with respect to  $B$ . The Casimir element  $C$  for  $B$  is the element of the universal enveloping algebra  $U(\mathfrak{g})$  given by the formula

$$C = \sum_{i=1}^n X_i X^i. \quad (14.54)$$

Although the definition relies on a choice of basis for the Lie algebra, it is easy to show that  $C$  is independent of this choice. On the other hand,  $C$  does depend on the bilinear form  $B$ . The invariance of  $B$  implies that the Casimir element commutes with all elements of the Lie algebra  $\mathfrak{g}$ , and hence lies in the center of the universal enveloping algebra  $U(\mathfrak{g})$ .

In other words, it is a member of the algebra of all differential operators that commutes with all the generators in the Lie algebra. In fact all quadratic elements in the center of the universal enveloping algebra arise this way. However, the center may contain other, non-quadratic, elements.

The Casimir operator must explicitly commute with the Lie bracket, i.e. one must have that

$$[C_{(m)}, X_i] = 0 \quad (14.55)$$

for all basis elements  $X_i$  of the algebra.

For a semi-simple Lie algebra the dimension of the center of the universal enveloping algebra is equal to its rank, so number of Casimir operators of a Lie algebra is equal to its rank.

A prototypical example is the Casimir invariant is squared angular momentum operator, which is a Casimir invariant of the three-dimensional rotation group. The Lie algebra of  $SO(3)$ , the rotation group for three-dimensional Euclidean space, is simple of rank 1. So it has a single independent Casimir, which is simply the sum of the squares of the generators  $L_x, L_y, L_z$  of the algebra.

That is, the Casimir invariant is given by

$$L^2 = L_x^2 + L_y^2 + L_z^2. \quad (14.56)$$

The invariance of the Casimir operator implies that it is a multiple of the identity operator  $I$ . This constant can be computed explicitly, giving the following result

$$L^2 = L_x^2 + L_y^2 + L_z^2 = \ell(\ell + 1)I . \quad (14.57)$$

In quantum mechanics, the scalar value  $\ell$  is referred to as the total angular momentum. For finite-dimensional matrix-valued representations of the rotation group,  $\ell$  always takes on integer values (for bosonic representations) or half-integer values (for fermionic representations).

Another important example is the rank-2 Poincaré group, which has two Casimir operators – one second order (Klein-Gordon operator) and one fourth order (square of the Pauli-Lubanski vector, see the Exercise: 7.3). Eigenvalues of these operators correspond to the particle mass and spin, respectively.

**Exercise 14.1:** Work out the symmetry group of a square. How many elements does it have? Construct the multiplication table, and determine whether or not the group is Abelian.

**Exercise 14.2:** Show that any  $n \times n$  unitary matrix  $U^\dagger U = 1$  can be written as  $U = e^{iH}$ , where  $H$  is hermitian,  $H^\dagger = H$ , and that  $\det U = 1$  implies that  $H$  is traceless.

**Exercise 14.3:** Show that  $SU(n)$  group has  $n^2 - 1$  independent group parameters and the maximum number of mutually commuting matrices is  $(n - 1)$ .



## Chapter 15

# Examples of Lie Groups

Before getting back to physics, we give a spoiler of how Lie theory is used in physics. What we are going to find is that some physical interaction (electromagnetism, weak force, strong force) will ultimately be described by a Lie group in some particular representation.

The particles that interact with that force will be described by the eigenvectors of the Cartan generators of the group, and the eigenvalues of those eigenvectors will be the physically measurable charges. Clearly, the number of charges associated with the interaction is equal to the number of dimensions of the representation. For example, you likely are aware that the strong force has 3 charges, called 'colors' (red, green, and blue). So, the strong force (we will see) will be in a 3-dimensional representation of the group that describes it. We will find that all forces carrying particles (photons, gluons,  $W$  and  $Z$  bosons) will be described by the generators of their respective Lie group. The Cartan generators will be force-carrying particles which can interact with any particle charged under that group by transferring energy and momentum, but do not change the charge (photons and  $Z$  bosons). This makes sense because Cartan generators are not raising or lowering operators. On the other hand, the non-Cartan generators will be force carrying particles which interact with any particle charged under that group by not only transferring energy and momentum, but also changing the charge ( $W$  bosons and gluons).

We won't be able to come back to discussing how this works until some examples are worked out, otherwise this may not be clear.

### 15.1 $SO(2)$

We start by looking at an extremely simple group,  $SO(2)$ . This is the group of rotations in the plane that leaves

$$\bar{r}^2 = x^2 + y^2 = (x \ y) \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \bar{v}^T \cdot \bar{v} \quad (15.1)$$

invariant. So for some generator  $X$  (which we will now find) of  $SO(2)$ ,

$$\bar{v} \rightarrow R(\theta)\bar{v} = e^{i\theta X}\bar{v}, \quad \bar{v}^T \rightarrow \bar{v}^T e^{i\theta X^T}. \quad (15.2)$$

So, expanding to first order only,

$$\bar{v}^T e^{i\theta X^T} e^{i\theta X}\bar{v} = \bar{v}^T(1 + i\theta X^T + i\theta X)\bar{v} = \bar{v}^T \cdot \bar{v} + \bar{v}^T i\theta(X + X^T)\bar{v}. \quad (15.3)$$

And because we demand that  $r^2$  be invariant, we demand that

$$X + X^T = 0 \Rightarrow X = -X^T. \quad (15.4)$$

So,  $X$  must be antisymmetric. Therefore we take

$$X \equiv \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

(the  $1/2$  is included to balance the  $i$  we inserted in equation (14.9) to ensure that  $X$  is Hermitian).

So, an arbitrary element of  $SO(2)$  will be

$$\begin{aligned} e^{i\theta X} &= e^{i\theta \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}} = e^{\theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}} = \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^0 + \theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^1 + \frac{1}{2}\theta^2 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 + \dots = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} - \frac{1}{2}\theta^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{3!}\theta^3 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \dots = \\ &= \begin{pmatrix} 1 - \frac{1}{2}\theta^2 + \dots & \theta - \frac{1}{3!}\theta^3 + \dots \\ -(\theta - \frac{1}{3!}\theta^3 + \dots) & 1 - \frac{1}{2}\theta^2 + \dots \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \end{aligned} \quad (15.5)$$

which is exactly what we would expect for a matrix describing rotations in the plane.

Also, notice that because  $SO(2)$  is Abelian, the commutation relations trivially vanish ( $[X, X] \equiv 0$ ), and so all of the structure constants are zero.

Now that we have found an explicit example of a generator, and seen an example of how generators relate to group elements, we move on to slightly more complicated examples.

## 15.2 $SO(3)$

We could easily generalize the argument from the preceding section and find the generators of  $SO(3)$  in the same way, but in order to illustrate more clearly how generators work, we will approach  $SO(3)$  differently by working backwards. Above, we found the generators and used them to calculate the group elements. Here, we begin with the known group elements of  $SO(3)$ , which are just the standard Euler matrices for rotations in 3-dimensional space:

$$R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{pmatrix}, \quad (15.6)$$

$$R_y(\psi) = \begin{pmatrix} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ \sin \psi & 0 & \cos \psi \end{pmatrix}, \quad (15.7)$$

$$R_z(\theta) = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (15.8)$$

Now, recall the definition of the generators, equation (14.9). We can use it to find the generators of  $SO(3)$ , which we will denote  $J_x$ ,  $J_y$ , and  $J_z$ .

$$\begin{aligned} J_x &= \frac{1}{i} \left. \frac{dR_x(\phi)}{d\phi} \right|_{\phi=0} = \frac{1}{i} \left. \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin \phi & \cos \phi \\ 0 & -\cos \phi & \sin \phi \end{pmatrix} \right|_{\phi=0} = \\ &= \frac{1}{i} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \end{aligned} \quad (15.9)$$

And similarly

$$J_y = \frac{1}{i} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad J_z = \frac{1}{i} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (15.10)$$

You can plug these into the exponentials with the appropriate parameters ( $\phi$ ,  $\psi$ , or  $\theta$ ) and find that  $e^{i\phi J_x}$ ,  $e^{i\psi J_y}$ , and  $e^{i\theta J_z}$  reproduce (15.6), (15.7), and (15.8), respectively.

Furthermore, you can multiply out the commutators to find

$$[J_x, J_y] = iJ_z, \quad [J_y, J_z] = iJ_x, \quad [J_z, J_x] = iJ_y, \quad (15.11)$$

or

$$[J_i, J_j] = i\epsilon_{ijk}J_k, \quad (15.12)$$

which tells us that the structure constants for  $SO(3)$  are

$$f_{ijk} = \epsilon_{ijk}, \quad (15.13)$$

where  $\epsilon_{ijk}$  is the totally antisymmetric tensor. The structure constants being non-zero is consistent with  $SO(3)$  being a non-Abelian group.

### 15.3 $SU(2)$

We will approach  $SU(2)$  yet another way: by starting with the structure constants. It turns out they are the same as the structure constants for  $SO(3)$ :

$$f_{ijk} = \epsilon_{ijk}. \quad (15.14)$$

To see why, recall that  $SU(2)$  are rotations in two complex dimensions. The most general form of such a matrix  $U \in SU(2)$  is

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (15.15)$$

The 'Special' part of  $SU(2)$  demands that the determinant be equal to 1, or

$$ad - bc = 1, \quad (15.16)$$

and the 'Unitary' part demands that

$$U^{-1} = U^\dagger . \quad (15.17)$$

So,

$$U^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = U^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} , \quad (15.18)$$

or in other words,

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} , \quad (15.19)$$

where we demand

$$|a|^2 + |b|^2 = 1 . \quad (15.20)$$

Both  $a$  and  $b$  are in  $\mathbb{C}$ , and therefore have 2 real components each, so  $U$  has 4 real parameters. The constraint (15.20) fixes one of them, leaving 3 real parameters, just like in  $SO(3)$ . This is a loose explanation of why  $SU(2)$  and  $SO(3)$  have the same structure constants. They are both rotational groups with 3 real parameters.

This also tells us that  $SU(2)$  will have 3 generators.

### 15.3.1 $SU(2)$ and Physical States

The elements of any Lie group (in a  $d$ -dimensional representation consisting of  $d \times d$  matrices) will act on vectors, just like the  $3 \times 3$  matrices representing  $S_3$  acted on  $(R \ O \ Y)^T$  in section 13.4. The most natural way to understand the space a Lie group acts on is to study the eigenvectors and eigenvalues of the generators of the representation you are using (the reason for this is beyond the scope of these notes at this point, but will become more clear as we proceed). These eigenvectors will obviously form a basis of the eigenspace of the physical space the group is acting on.

Using similarity transformations, one or more of the generators of a Lie group can be diagonalized. For now, trust us that with  $SU(2)$ , it is only possible to diagonalize one of the three generators at a time (you may convince yourself of this by studying the commutation relations). We will call the generators of  $SU(2)$   $J^1$ ,  $J^2$ , and  $J^3$ , and by convention we take  $J^3$  to be the diagonal one. So, consequently, the eigenvectors of  $J^3$  will be the basis vectors of the physical vector space upon which  $SU(2)$  acts.

Now, we know that  $J^3$  (whatever it is ... we don't know at this point) will in general have more than one eigenvalue. Let's call the greatest eigenvalue of  $J^3$  (whatever it is)  $j$ , and the eigenvectors of  $J^3$  will be denoted  $|j; m\rangle$  (the first  $j$  is merely a label – the second value describes the vector), where  $m$  is the eigenvalue of the eigenvector. The eigenvector corresponding to the greatest eigenvalue  $j$  will obviously then be  $|j; j\rangle$ . So,

$$J^3|j; j\rangle = j|j; j\rangle, \quad (15.21)$$

or more generally

$$J^3|j; m\rangle = m|j; m\rangle. \quad (15.22)$$

Now let's assume that we know  $|j; j\rangle$ . There is a trick we can employ to find the rest of the states. Define the following linear combinations of the generators:

$$J^\pm \equiv \frac{1}{\sqrt{2}}(J^1 \pm iJ^2). \quad (15.23)$$

Now, using the fact that the  $SU(2)$  generators obey the commutation relations in equation (15.13), it is easy to show the following relations,

$$[J^2, J^\pm] = \pm J^\pm \quad \text{and} \quad [J^+, J^-] = J^3. \quad (15.24)$$

Notice that, because by definition  $J^i$  are all Hermitian, we have

$$(J^-)^\dagger = J^+. \quad (15.25)$$

Consider some arbitrary eigenvector  $|j; m\rangle$ . We know from (15.22) that the eigenvalue of this will be  $m$ . But now let's create some new state by acting on  $|j; m\rangle$  with either of the operators (15.23). The new state will be  $J^\pm|j; m\rangle$ , but what will the  $J^3$  eigenvalue be? Using the commutation relations in (15.24),

$$J^3 J^\pm|j; m\rangle = (\pm J^\pm + J^\pm J^3)|j; m\rangle = (m \pm 1)J^\pm|j; m\rangle. \quad (15.26)$$

So, the vector  $J^+|j; m\rangle$  is the eigenvector with eigenvalue  $m + 1$ , and the vector  $J^-|j; m\rangle$  is the eigenvector with eigenvalue  $m - 1$ .

If we have some arbitrary eigenvector  $|j; m\rangle$ , we can use  $J^\pm$  to move up or down to the eigenvector with the next highest or lowest eigenvalue. For this reason,  $J^\pm$  are called the *Raising* and *Lowering* operators. They raise and lower the eigenvalue of the state by one.

Clearly, the eigenvector with the greatest eigenvalue  $j$ , with eigenvector  $|j; j\rangle$ , cannot be raised any higher, so we define

$$J^+|j; j\rangle \equiv 0, \quad (15.27)$$

We will see that there is also a lowest eigenvalue  $j'$ , so we similarly define

$$J^-|j; j'\rangle \equiv 0. \quad (15.28)$$

Now, considering once again  $|j; j\rangle$ . We know that if we operate on this state with  $J^-$ , we will get the eigenvector with the eigenvalue  $j-1$ . But, we don't know exactly what this state will be (knowing the eigenvalue doesn't mean we know the actual state). But, we know it will be proportional to  $|j; j-1\rangle$ . So, we set

$$J^-|j; j\rangle = N_j|j; j-1\rangle, \quad (15.29)$$

where  $N_j$  is the proportionality constant. To find  $N_j$ , we take the inner product (and using (15.25)):

$$\langle j; j|J^+J^-|j; j\rangle = |N_j|^2\langle j; j-1|j; j-1\rangle. \quad (15.30)$$

But we can also write

$$\begin{aligned} \langle j; j|J^+J^-|j; j\rangle &= \langle j; j|(J^+J^- - J^-J^+)|j; j\rangle = \\ &= \langle j; j|[J^+, J^-]|j; j\rangle = \\ &= \langle j; j|J^3|j; j\rangle = j\langle j; j|j; j\rangle = j, \end{aligned} \quad (15.31)$$

where we used (15.27) to get the first equality, and (15.24) to get the third equality. We also assumed that  $|j; j\rangle$  is normalized.

So, (15.31) tells us

$$\langle j; j-1|j; j-1\rangle = 1 \iff N_j \equiv \sqrt{j}. \quad (15.32)$$

And our normalized state is therefore

$$\frac{J^-}{N_j}|j; j\rangle = \frac{J^-}{\sqrt{j}}|j; j\rangle = |j; j-1\rangle. \quad (15.33)$$

Repeating this to find  $N_{j-1}$ , we have

$$\begin{aligned}
|N_{j-1}|^2 &= \langle j; j-2 | j; j-2 \rangle = \langle j; j-1 | J^+ J^- | j; j-1 \rangle = \\
&= \left\langle j; j \left| \frac{J^+}{\sqrt{j}} J^+ J^- \frac{J^-}{\sqrt{j}} \right| j; j \right\rangle = \frac{1}{j} \langle j; j | J^+ J^+ J^- J^- | j; j \rangle = \\
&= \frac{1}{j} \langle j; j | J^+ (J^3 + J^- J^+) J^- | j; j \rangle = \\
&= \frac{1}{j} \langle j; j | (J^+ J^3 J^- + J^+ J^- J^+ J^-) | j; j \rangle = \tag{15.34} \\
&= \frac{1}{j} \langle j; j | (J^+ (-J^- + J^- J^3) + J^+ J^- (J^3 + J^- J^+)) | j; j \rangle = \\
&= \frac{1}{j} [\langle j; j | (-J^+ J^- + j J^+ J^- + j J^+ J^-) | j; j \rangle] = \\
&= \frac{1}{j} \langle j; j | (-[J^+, J^-] + 2j[J^+, J^-]) | j; j \rangle = \\
&= \frac{1}{j} \langle j; j | (-J^3 + 2jJ^3) | j; j \rangle = \frac{1}{j} (2j^2 - j) = 2j - 1 .
\end{aligned}$$

So,

$$N_{j-1} = \sqrt{2j-1} . \tag{15.35}$$

We can continue this process, and we will find that the general result is

$$N_{j-k} = \frac{1}{\sqrt{2}} \sqrt{(2j-k)(k+1)} , \tag{15.36}$$

and the general states are defined by

$$|j; j-k\rangle = \frac{1}{N_{j-k}} (J^-)^k |j; j\rangle . \tag{15.37}$$

Notice that these expressions recover (15.32) and (15.34) for  $k=0$  and  $k=1$ , respectively.

Furthermore, notice that when  $k=2j$ ,

$$N_{j-2j} = \frac{1}{\sqrt{2}} \sqrt{(2j-2j)(2j+1)} \equiv 0 . \tag{15.38}$$

So, the state  $|j; j-k\rangle|_{k=2j} = |j; -j\rangle$  is the state with the lowest eigenvalue, and by definition  $J^- |j; -j\rangle \equiv 0$ .



So, in a general representation of  $SU(2)$ , we have  $2j + 1$  states:

$$\{j, j-1, j-2, \dots, -j+2, -j+1, -j\}. \quad (15.39)$$

This therefore demands that  $j = n/2$  for some integer  $n$ . In other words, the highest eigenvalue of an  $SU(2)$  eigenvector can be  $0, 1/2, 1, 3/2, 2$ , etc.

Furthermore, using these states, it is easy to show

$$\begin{aligned} \langle j; m' | J^3 | j; m \rangle &= m \delta_{m', m}, \\ \langle j; m' | J^+ | j; m \rangle &= \frac{1}{\sqrt{2}} \sqrt{(j+m+1)(j-m)} \delta_{m', m+1}, \\ \langle j; m' | J^- | j; m \rangle &= \frac{1}{\sqrt{2}} \sqrt{(j+m)(j-m+1)} \delta_{m', m-1}. \end{aligned} \quad (15.40)$$

### 15.3.2 $SU(2)$ for $j = 1/2$

We will skip the  $j = 0$  case because it is trivial (though we will discuss it later when we return to physics).

For  $j = 1/2$ , the two eigenvalues of  $J^3$  will be  $1/2$  and  $1/2 - 1 = -1/2$ . So, denoting the  $J^3$  generator of  $SU(2)$  when  $j = 1/2$  as  $J_{1/2}^3$ , we have

$$J_{1/2}^3 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}. \quad (15.41)$$

Now, inverting (15.23) to get

$$J^1 = \frac{1}{\sqrt{2}}(J^- + J^+) \quad \text{and} \quad J^2 = \frac{i}{\sqrt{2}}(J^- - J^+), \quad (15.42)$$

and using the standard matrix equation  $[J_j^a]_{m', m} = \langle j, m' | J^a | j, m \rangle$ , and the explicit products in (15.40), we can find (for example)

$$\left\langle \frac{1}{2}; -\frac{1}{2} \left| J^1 \right| \frac{1}{2}; -\frac{1}{2} \right\rangle = \left\langle \frac{1}{2}; -\frac{1}{2} \left| \frac{1}{\sqrt{2}}(J^- + J^+) \right| \frac{1}{2}; -\frac{1}{2} \right\rangle = \dots = 0. \quad (15.43)$$

So  $[J^1]_{11} = 0$ . Then,

$$\left\langle \frac{1}{2}; -\frac{1}{2} \left| J^1 \right| \frac{1}{2}; \frac{1}{2} \right\rangle = \left\langle \frac{1}{2}; -\frac{1}{2} \left| \frac{1}{\sqrt{2}}(J^- + J^+) \right| \frac{1}{2}; \frac{1}{2} \right\rangle = \dots = \frac{1}{2}. \quad (15.44)$$

So  $[J^1]_{12} = 1/2$ .

We can continue this to find all the elements for each generator for  $j = 1/2$ . The final result will be

$$\begin{aligned} J_{1/2}^1 &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\sigma^1}{2}, \\ J_{1/2}^2 &= \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\sigma^2}{2}, \\ J_{1/2}^3 &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\sigma^3}{2}, \end{aligned} \tag{15.45}$$

where the  $\sigma^i$  matrices are the *Pauli Spin Matrices*.

This is no accident! We will discuss this in more detail later, but for now recall that we said that  $SU(2)$  is the group of transformations in 2-dimensional complex space (with one of the real parameters fixed, leaving 3 real parameters).

We are going to see that  $SU(2)$  is the group which represents quantum mechanical spin, where  $j$  is the value of the spin of the particle. In other words, particles with spin 1/2 are described by the  $j = 1/2$  representation (the  $2 \times 2$  representation in (15.45)), and particles with spin 1 are described by the  $j = 1$  representation, and so on. In other words,  $SU(2)$  describes quantum mechanical spin in 3 dimensions in the same way that  $SO(3)$  describes normal 'spin' in 3 dimensions. We will talk about the physical implications, reasons, and meaning of this later. However, as a warning, be careful at this point not to think too much in terms of physics.

You have likely covered  $SU(2)$  in great detail in a quantum mechanics course (though you may not have known it was called ' $SU(2)$ '), but the approach we are taking here has a different goal than what you have likely seen before.

The properties of  $SU(2)$  we are seeing here are actually very specific and simplified illustrations of much deeper concepts in Lie groups, and in order to understand particle physics we must understand Lie groups in this way. So for now, try to fight the temptation to merely understand everything we are doing in terms of the physics you have seen before and learn this as we are presenting it: pure mathematics. We will focus on how it applies to physics later, in its fuller and more fundamental way than introductory quantum mechanics makes apparent.

**15.3.3  $SU(2)$  for  $j = 1$** 

You can follow the same procedure we used above to find

$$\begin{aligned} J_1^1 &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ J_1^2 &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \\ J_1^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \end{aligned} \tag{15.46}$$

Notice that only  $J_1^3$  is diagonal (as before), and that the eigenvalues are  $\{1, 0, -1\}$ , or  $\{j, j-1, j-2 = -j\}$  as we'd expect.

**15.3.4  $SU(2)$  for Arbitrary  $j$** 

For any given  $j$ , we have 3 generators  $J_j^1$ ,  $J_j^2$ , and  $J_j^3$ , and for whatever dimension ( $d = 2j + 1$ ) the physical space we are working in, we have  $d$  eigenvectors

$$|j; j\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |j; j-1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad |j; -j\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \tag{15.47}$$

with eigenvalues  $\{j, j-1, j-2, \dots, -j\}$ , respectively.

Then, for any  $j$ , we can form the linear combinations

$$J_j^\pm \equiv \frac{1}{\sqrt{2}}(J_j^1 \pm iJ_j^2). \tag{15.48}$$

For example, for  $j = 1/2$  these are

$$J_{1/2}^+ = \frac{1}{\sqrt{2}} \left[ \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{i}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right] = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{15.49}$$

and similarly

$$J_{1/2}^- = \dots = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (15.50)$$

So, the two  $j = 1/2$  eigenvectors will be

$$\left| \frac{1}{2}; \frac{1}{2} \right\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \left| \frac{1}{2}; -\frac{1}{2} \right\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (15.51)$$

So,

$$\begin{aligned} J_{1/2}^+ \left| \frac{1}{2}; \frac{1}{2} \right\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0, \\ J_{1/2}^- \left| \frac{1}{2}; -\frac{1}{2} \right\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0, \end{aligned} \quad (15.52)$$

and similarly

$$J_{1/2}^- \left| \frac{1}{2}; \frac{1}{2} \right\rangle = \left| \frac{1}{2}; -\frac{1}{2} \right\rangle, \quad J_{1/2}^+ \left| \frac{1}{2}; -\frac{1}{2} \right\rangle = \left| \frac{1}{2}; \frac{1}{2} \right\rangle, \quad (15.53)$$

which is exactly what we would expect.

The same calculation can be done for the  $j = 1$  case and we will find the same results, except that the  $j = 1$  state (the first eigenvector) can be lowered twice. The first time  $J_{1/2}^-$  acts it takes it to the state with eigenvalue 0, and the second time it acts it takes it to the state with eigenvalue  $-1$ . Acting a third time will destroy the state (take it to 0). Analogously, the lowest state, with eigenvalue  $j = -1$  can be raised twice.

We can do the same analysis for any  $j = \text{integer or half integer}$ .

As we said before, we interpret  $j$  as the quantum mechanical spin of a particle, and the group  $SU(2)$  describes that rotation. It is important to recognize that quantum spin is not a rotation through spacetime (it would be described by  $SO(3)$  if it was), but rather through the mathematically constructed spinor space. We will talk more about this space later.

So for a given particle with spin, we can talk about both its rotation through physical spacetime using  $SO(3)$ , as well as its rotation through complex spinor space using  $SU(2)$ . Both values will be physically measurable and will be conserved quantities. The total angular momentum of the particle

will be the combination of both spin and spacetime angular momentum. Again, we will talk much more about the spin of physical particles when we return to a discussion of physics. We only mention this now to give a preview of where this is going. However, spin is not the only thing  $SU(2)$  describes. We will also find that it is the group which governs the weak nuclear force (whereas  $U(1)$  describes the electromagnetic force, and  $SU(3)$  describes the strong force ... much, much more on this later).

### 15.3.5 Adjoint Representation of $SU(2)$

We now illustrate what we did in section 14.5 with  $SU(2)$ . We will work in the adjoint representation to make the correspondence with section 14.5 as transparent as possible.  $SU(2)$  has 3 generators, and therefore the adjoint representation will consist of  $3 \times 3$  matrices. This is simply the  $j = 1$  representation, which we wrote out in equation (15.46).

First, it is easy to verify that (14.28) and (14.32) hold for  $\kappa = 2$ . Next we look at the eigenstates. We know they will be the normal vectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (15.54)$$

(we will relabel them to be consistent with section 14.5 shortly).

Obviously only  $J_1^3$  is diagonal, so  $SU(2)$  has rank  $M = 1$ . We define

$$\begin{aligned} H^1 = J_1^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \\ E^1 = J_1^1 &= \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ E^2 = J_1^2 &= \frac{1}{2} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \end{aligned} \quad (15.55)$$

Because the rank is 1, the root vectors will be 1-dimensional vectors, or scalars. We find them easily by finding the eigenvalues of each eigenvector

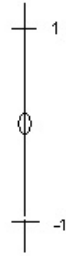
with  $H^1$ :

$$H^1 v_1 = (+1)v_1, \quad H^1 v_2 = (0)v_2, \quad H^1 v_3 = (-1)v_3. \quad (15.56)$$

So the root vectors are

$$\bar{t}_1 = t_1 = +1, \quad \bar{t}_2 = t_2 = 0, \quad \bar{t}_3 = t_3 = -1. \quad (15.57)$$

We can graph these on the real line as shown below,



Now our initial guess will be to associate  $v_3$  with  $J_1^3 = H^1$ , and then  $v_1 = E^1$  and  $v_2 = E^2$ . But we want to exploit what we learned in section 14.5, and therefore we must make sure that (14.43) and (14.52) hold.

Starting with (14.43), we check (leaving the tedious matrix multiplication up to you)

$$[H^1, E^1] = \dots = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad (15.58)$$

$$[H^1, E^2] = \dots = -\frac{i}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (15.59)$$

But we have a problem. According to (14.43),  $[H^1, E^i]$  should be proportional to  $E^i$ , but this is not the case here. However notice that in (15.58),

$$\frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = iE^2, \quad (15.60)$$

and in (15.59),

$$-\frac{i}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = -iE^1 . \quad (15.61)$$

Writing this more suggestively,

$$[H^1, E^1] = iE^2 , \quad [H^1, iE^2] = E^1 . \quad (15.62)$$

So, if we take the linear combinations of equations (15.62), we get

$$[H^1, \alpha E^1 \pm \beta iE^2] = \beta E^1 \pm \alpha iE^2 , \quad (15.63)$$

which has the correct form of equation (14.43) as long as  $\alpha = \beta$ . Therefore we are now working with the operators

$$E^\pm \equiv \alpha(E^1 \pm iE^2) . \quad (15.64)$$

Now we seek to impose (14.52). We start by evaluating

$$\begin{aligned} [E^+, E^-] &= \alpha^2 [E^1 + iE^2, E^1 - iE^2] = \\ &= \alpha^2 ([E^1, E^1] - i[E^1, E^2] + i[E^2, E^1] + [E^2, E^2]) = \\ &= -2i\alpha^2 [E^1, E^2] = \dots = -2i\alpha^2 iH^1 = 2\alpha^2 H^1 , \end{aligned} \quad (15.65)$$

Then, from equations (15.57) and the definition of  $E^\pm$ , we see that

$$\pm e_1^1 = \pm(t_1 - t_2) = \pm(1 - 0) = \pm 1 . \quad (15.66)$$

So we therefore set

$$\alpha^2 = \frac{1}{2} \Rightarrow \alpha = \frac{1}{\sqrt{2}} , \quad (15.67)$$

and we find that the appropriate non-Cartan generators (including the 1 to be consistent with the notation in section 14.5) are

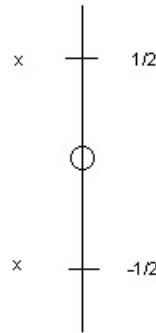
$$E^{\pm 1} = \frac{1}{\sqrt{2}}(E^1 \pm iE^2) \quad (15.68)$$

which is exactly what we had in equation (15.23) above. So, we have derived the trick used to understand quantum mechanical spin in introductory quantum courses!

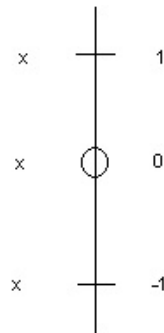
### 15.3.6 $SU(2)$ for Arbitrary $j \dots$ Again

Now that we have our operators in the adjoint representation, we can consider any arbitrary representation. As we saw in section 15.3.4, we can form the linear combinations in equation (15.68) for any  $j = \text{integer or half integer}$ . The weight vectors will always look like displayed those in the diagram on page 228 (in other words, raising and lowering operators always raise or lower their eigenvalue by 1).

The space of physical states, on the other hand, changes for each representation. For  $j = \frac{1}{2}$ , we have

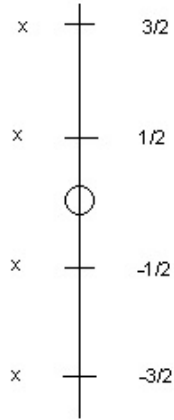


For  $j = 1$ ,



For  $j = 3/2$ ,





and so on.

Notice that the vectors graphed in the diagram on the page 228 are the exact vectors required to move from point to point in each of these graphs. This is obviously not a coincidence.

## 15.4 $SU(3)$

Now that we have said pretty much everything we can about  $SU(2)$ , which is only Rank 1 (and therefore not all that interesting), we move on to  $SU(3)$ . However, we will expedite the process by stating the structure constants up front. The non-zero structure constants are:

$$\begin{aligned}
 f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} &= \frac{1}{2}, \\
 f_{458} = f_{678} &= \frac{\sqrt{3}}{2}, \\
 f_{123} &= 1.
 \end{aligned}
 \tag{15.69}$$

The most convenient representation is the *Fundamental Representation* (consisting of  $3 \times 3$  matrices). They are

$$T^a = \frac{1}{2} \lambda^a, \quad (a = 1, \dots, 8)
 \tag{15.70}$$

where

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & (15.71) \\ \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned}$$

Only two of these are diagonal,  $\lambda^3$  and  $\lambda^8$ . So,  $SU(3)$  is a rank 2 group.

Working in the adjoint representation of  $SU(3)$  would involve  $8 \times 8$  matrices, which would obviously be very tedious. So, we exploit the fact that the techniques we developed in section 14.5 are valid in any representation, and stick with the Fundamental Representation defined by the generators in (15.71).

Proceeding as in section 15.3.5, we note that the eigenvectors will again be

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (15.72)$$

Then, the Cartan generators are

$$H^1 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad H^2 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad (15.73)$$

and the non-Cartan Generators are simply

$$E^1 = T^1, \quad E^2 = T^2, \quad E^3 = T^4, \quad E^4 = T^5, \quad E^5 = T^6, \quad E^6 = T^7. \quad (15.74)$$

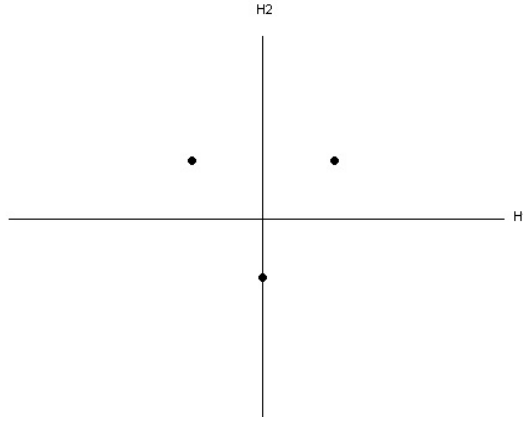
So we have 6 eigenvalues to find,

$$\begin{aligned} H^1 v_1 &= \left(\frac{1}{2}\right) v_1, & H^1 v_2 &= \left(-\frac{1}{2}\right) v_2, & H^1 v_3 &= (0) v_3, \\ H^2 v_1 &= \left(\frac{1}{2\sqrt{3}}\right) v_1, & H^2 v_2 &= \left(\frac{1}{2\sqrt{3}}\right) v_2, & H^2 v_3 &= \left(-\frac{1}{\sqrt{3}}\right) v_3. \end{aligned} \quad (15.75)$$

The weight vectors will be 2-dimensional (because the rank is 2):

$$\bar{t}^1 = \left(\frac{1}{2} \quad \frac{1}{2\sqrt{3}}\right)^T, \quad \bar{t}^2 = \left(-\frac{1}{2} \quad \frac{1}{2\sqrt{3}}\right)^T, \quad \bar{t}^3 = \left(0 \quad -\frac{1}{\sqrt{3}}\right)^T. \quad (15.76)$$

We can graph these in  $\mathbb{R}^2$  as shown below,



Repeating nearly the identical argument we started with equation (15.58) and repeating it for all 6 non-Cartan generators, we find that in order to maintain (14.43) and (14.52), we must work with the operators

$$\begin{aligned} \frac{1}{\sqrt{2}}(T^1 \pm iT^2) &= \frac{1}{\sqrt{2}}(E^1 \pm iE^2), \\ \frac{1}{\sqrt{2}}(T^4 \pm iT^5) &= \frac{1}{\sqrt{2}}(E^3 \pm iE^4), \\ \frac{1}{\sqrt{2}}(T^6 \pm iT^7) &= \frac{1}{\sqrt{2}}(E^5 \pm iE^6). \end{aligned} \quad (15.77)$$

The weight vectors associated with these will be, respectively,

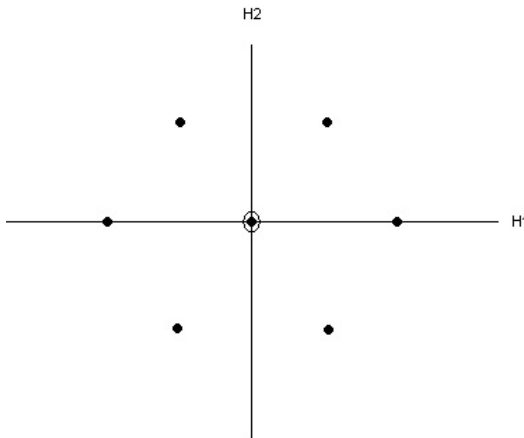
$$\pm(\bar{t}_1 - \bar{t}_2) = \pm \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \pm(\bar{t}_1 - \bar{t}_3) = \pm \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}, \quad \pm(\bar{t}_2 - \bar{t}_3) = \pm \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}.$$

So, the non-Cartan generators are

$$E^{\pm} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad E^{\pm} \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}, \quad E^{\pm} \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}. \quad (15.78)$$

We are no longer in the adjoint representation. What we did here is more general; we chose them to be the differences in the three weight vectors in equation (15.76), so that these vectors would naturally transform from one eigenvector to another (just as the raising and lowering operators do, as we found for  $SU(2)$  and more generally in section 14.5). The remarkable property of Lie groups is that this is always possible in any representation.

We can graph the 6 vectors in (15.78), along with the two Cartan weight vectors, which we know from (14.39) are 0:



And again, just as with  $SU(2)$ , notice that the 6 non-zero vectors are the exact vectors that would be necessary to move from point to point on the above diagram. So once again, we see that the non-Cartan generators act as raising and lowering operators which transform between the eigenstates of the Cartan generators. Notice that there were 6 non-Cartan generators, and they formed linear combinations to form 6 raising and lowering operators.

**Exercise 15.1:** The third Pauli matrix  $\sigma_3$  is in the group  $O(2)$  or in  $SO(2)$ ?

**Exercise 15.2:** Suppose  $\psi_1$  and  $\psi_2$  are the bases for the spin-1/2 representation of  $SU(2)$  having eigenvalues of  $\pm 1/2$  for the diagonal generator  $T_3$ ,  $T_3\psi_1 = \psi_1/2$  and  $T_3\psi_2 = -\psi_2/2$ . Calculate the eigenvalues of  $T_3$  operating on  $\psi_1^*$  and  $\psi_2^*$ , respectively.

**Exercise 15.3:** The  $\rho$  vector meson has isospin 1 (it has three charge states:  $\rho^+$ ,  $\rho^0$ ,  $\rho^-$ ). Construct the  $SU(2)$  invariant  $\rho\pi\pi$  coupling.

## Part VI

# Lecture – Particle Experiments



# Chapter 16

## Anatomy of Experiments

In this lecture we will try to describe the main landscape of experimental high energy physics today.

### 16.1 Accelerators and Laboratories

Currently the only possibility to investigate properties of elementary particles is to study their collisions. In general, the heavier the particle you want to produce, the higher must be the energy of the collision. In quantum-mechanical terms, a particle of momentum  $p$  has an associated wavelength  $\lambda$  given by the de Broglie formula  $\lambda = h/p$ , where  $h$  is the Planck constant. At large wavelengths (low momenta) you can only hope to resolve relatively large structures; in order to examine something extremely small, you need comparably short wavelengths, and hence high momenta. If you like, consider this a manifestation of the uncertainty principle ( $\Delta x \Delta p \geq h/4\pi$ ) – to make  $\Delta x$  small,  $\Delta p$  must be large. The conclusion is the same: to probe small distances you need high energies.

A particle accelerator is a machine that uses electromagnetic fields to propel charged particles to nearly light speed (having high energies) and to contain them in well-defined beams. In experimental particle physics currently accelerators are shaped in one of two ways:

- *Linear Colliders*: An example of such an accelerator is the Stanford

Linear Accelerator Center (SLAC);

- *Circular or Synchrotron Accelerators:* These provide higher energies than a Linear ones, such as the Large Hadron Collider at CERN.

Accelerators can also be arranged to provide collisions of two types:

- *Fixed Target Experiments:* When particles are shot at a fixed target;
- *Colliding Beam Experiments:* When two beams of particles are made to cross each other.

### 16.1.1 Luminosity

Important parameters in colliders are the energy of the beams and the rate of collisions,  $R$ , or the luminosity  $L$ . The rate of collisions is defined as:

$$R = \frac{dN}{dt} = L\sigma , \quad (16.1)$$

where  $dN/dt$  is the number of hard collision events produced per second, and  $\sigma$  is the cross section of the process. Integrating over time, we get:

$$N_{EP} = \sigma \times \int Ldt , \quad (16.2)$$

where  $N_{EP}$  are the number of produced hard collision events (*Events Produced*) of the process with cross section  $\sigma$  and  $\int Ldt$  is the integrated luminosity which is provided by the accelerator in a given time period. Unfortunately, a given high energy physics detector does not observe every collision event that is produced. For example, the trigger is inefficient, as is the identification of the final state particles, and some fraction of the events may be produced beyond the detector acceptance. These inefficiencies need to be experimentally evaluated and once accounted for the expression becomes:

$$N_{EO} = \sigma \times \int Ldt \times \epsilon , \quad (16.3)$$

where  $N_{EO}$  is now the number of *Events Observed* in the detector, and  $\epsilon$  is the total efficiency of identifying the collision event of interest.



The units of a cross section are the same as the units of area and are typically represented by a *barn* ( $1 \text{ barn} = 10^{24} \text{ cm}^2$ ), for example,  $mb$ ;  $\nu b$ ;  $nb$ , etc. The units of *instantaneous luminosity* are the same as the units of  $[1/(\text{cross section} \times \text{time})]$ , for example  $\text{cm}^{-2}\text{s}^{-1}$ . *Integrated luminosity* has units of  $[1/\text{cross section}]$ , for example  $\text{cm}^{-2}$ , or  $pb^{-1}$ ,  $fb^{-1}$ , etc.

Next, let us consider an alternate expression for luminosity:

$$L = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \approx f \frac{n_b N^2}{4\pi\sigma_x\sigma_y} . \quad (16.4)$$

where  $n_1$  and  $n_2$  are the numbers of particles in each of the colliding bunches,  $f$  is the frequency with which they collide,  $\sigma_x$  and  $\sigma_y$  represent the size of the transverse beam (e.g. the RMS if we assume a Gaussian shaped beam),  $n_b$  is the number of bunches and  $N$  is the number of particles per bunch. So in order to increase the luminosity, it is important to squeeze as many colliding particles in as small a transverse beam spot as possible.

### 16.1.2 Particle Physics Laboratories

Nowadays the world's major particle physics laboratories are:

- *CERN* (European Organization for Nuclear Research) at Franco-Swiss border, near Geneva. Its main project is now the Large Hadron Collider (LHC), which had its first beam circulation on 10 September 2008, and is now the world's most energetic collider of protons. It also became the most energetic collider of heavy ions after it began colliding lead ions. Earlier facilities include the Large Electron-Positron Collider (LEP), which was stopped on 2 November 2000 and then dismantled to give way for LHC; and the Super Proton Synchrotron, which is being reused as a pre-accelerator for the LHC.
- *DESY* (Deutsches Elektronen-Synchrotron) in Hamburg, Germany. Its main facility is the Hadron Elektron Ring Anlage (HERA), which collides electrons and positrons with protons.
- *Brookhaven National Laboratory* (Long Island, US). Its main facility is the Relativistic Heavy Ion Collider (RHIC), which collides heavy ions such as gold ions and polarized protons. It is the world's first heavy ion collider and the world's only polarized proton collider.

- *Fermi National Accelerator Laboratory* (Fermilab) in Batavia, US. Its main facility until 2011 was the Tevatron, which collided protons and antiprotons and was the highest-energy particle collider on earth until the LHC surpassed it on 29 November 2009.
- *SLAC National Accelerator Laboratory* (Menlo Park, US). Its 2-mile-long linear particle accelerator began operating in 1962 and was the basis for numerous electron and positron collision experiments until 2008. Since then the linear accelerator is being used for the Linac Coherent Light Source X-ray laser as well as advanced accelerator design research. SLAC staff continue to participate in developing and building many particle detectors around the world.
- *Budker Institute of Nuclear Physics* (Novosibirsk, Russia). Its main projects are now the electron-positron colliders VEPP-2000, operated since 2006, and VEPP-4, started experiments in 1994. Earlier facilities include the first electron-electron beam-beam collider VEP-1, which conducted experiments from 1964 to 1968; the electron-positron colliders VEPP-2, operated from 1965 to 1974; and its successor VEPP-2M, performed experiments from 1974 to 2000.
- *Institute of High Energy Physics* (IHEP) in Beijing, China. IHEP manages a number of China's major particle physics facilities, including the Beijing Electron Positron Collider (BEPC), the Beijing Spectrometer (BES), the Beijing Synchrotron Radiation Facility (BSRF), the International Cosmic-Ray Observatory at Yangbajing in Tibet, the Daya Bay Reactor Neutrino Experiment, the China Spallation Neutron Source, the Hard X-ray Modulation Telescope (HXMT), the Accelerator-driven Sub-critical System (ADS) and the Jiangmen Underground Neutrino Observatory (JUNO).
- *KEK* (Tsukuba, Japan). It is the home of a number of experiments, such as the K2K (a neutrino oscillation experiment) and Belle (an experiment measuring the CP violation of  $B$  mesons).

### 16.1.3 What is After LHC?

A serious question confronting the experimental community is: what is the next machine after the LHC that we will want to build? Let us present some of the ideas and proposals being discussed.

The most serious option for a next machine after the LHC is an  $e^+e^-$  linear collider. This is a machine that would start with a CM energy of 500 GeV and might eventually be upgraded to 1.5 TeV. If the minimal supersymmetric SM is right, this machine could be a gold mine since most of the Higgses and superpartners would be accessible even at 500 GeV center of mass energy. While the discovery of SUSY might still take place at LHC, the  $e^+e^-$  linear collider offers a much cleaner experimental environment to fully study the SUSY particle spectrum. If, however, the minimal supersymmetric model is wrong, then this machine needs to push to the highest center of mass energies of over 1 TeV in order to extend our physics reach past what already will be learned from the LHC.

Another option that is being discussed as a future machine is an even higher energy proton-proton collider, with 100 TeV of energy in the CM. This would be a frontier machine that would push to the highest possible energy that is achievable by present technology.

A final option being discussed is a  $\mu^+\mu^-$  collider. One wins with muons over electrons since electron colliders are limited by radiation which is much suppressed for muons because of their larger mass. The talk is of a muon collider with 4 TeV in the CM.

## 16.2 Detectors

When the beams collide at an accelerator, physics happens: particles that we want to study emerge. The interaction region is instrumented with a detector that is designed to record as much information as possible about what is emerging from the beam collision.

The goal of every collider experiment is to completely surround the collision by arranging layers of different types of detectors. The form of the detector depends in its gross geometry on the accelerator type. At storage rings where the Lab frame is also the CM frame for the interaction, outgoing particles from the interaction are nearly isotropically distributed about the collision point and detectors reflect that fact. The detectors try to surround as much of the solid angle around the interaction point as possible, given practical and financial constraints.

It is not possible to describe a generic detector. Each detector is individu-

ally designed to match the machine at which it runs; however, all detectors are composed from a fairly consistent set of building blocks which can be easily described, although the execution or techniques used on different experiments will vary widely.

The particle detectors have the following general features, starting from center moving outwards:

- *Tracking Detectors within a Magnetic Field:* measures the charge, trajectory and momentum of charged particles;
- *Electromagnetic Calorimeter:* measures the energy and position of electromagnetic particles;
- *Hadronic Calorimeter:* measures the energy and position of hadronic particles;
- *Muon Chambers:* measures the trajectory and momentum (along with the tracking detector) of muons.

Also there are *Particle Identifications* of various sorts to distinguish different types of hadrons, particularly pions and kaons.

Below you can see the schematic drawings of the CMS (left) and ATLAS (right) detectors at CERN.



I will briefly discuss the various detector elements, mentioned above, and how they are most commonly used.

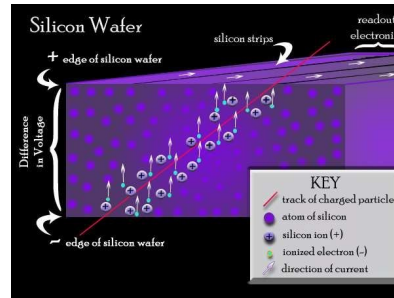
### 16.2.1 Tracking Detectors

A charged particle tracker, usually a drift chamber, is at the heart of most experiments. The main goal of tracking detectors is to measure the momentum, charge and trajectory of charged particles. Ideally, we want tracking detectors to contain as little material as possible in order to minimize multiple scattering.

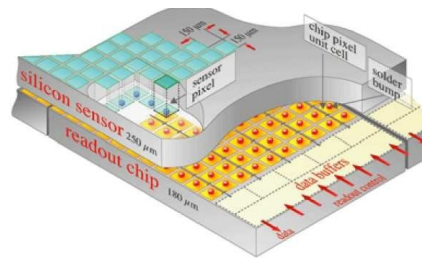
There are two main technologies of tracking detectors in particle physics:

- *Gas/Wire Drift Chambers:* These devices are made of wires in a volume filled with a gas, such as Argon-Ethane. They measure where a charged particle has crossed, when it ionizes the gas. There is an electrical potential applied to the wires, so atomic electrons knocked off the atoms in the gas drift to a positively charged sense wire. The chamber are connected to electronics which measure the charge of the signal and when it appears. To reconstruct the tracks of the charged particles several chamber planes are necessary. Advantages to drift chambers is their low thickness and are the traditionally preferred technology for large volume detectors. Typical single hit resolutions range from  $100 - 200 \mu m$ . An example of such a device is the CDF experiment's Central Outer Tracker (COT) which has approximately 30,000 wires. To measure the position of a track, a clock is started when the particle is produced (at the beam crossing) and stopped when the pulse height on a wire exceeds a preset value. Associated with each wire there will then be a time  $t_i$ . Using  $d_i = v_D t_i$  where  $v_D$  is the drift velocity of electrons in the gas, one can infer the distance from the wire to where the track ionization segment came from. By joining hits, one defines the track of the incident charged particle.
- *Silicon Detectors:* Precision silicon tracking devices work on the same physics principle as gas chambers, although the anode and cathode in a silicon detector are no longer wires but electrodes etched on a thin silicon wafer. Silicon detectors are usually placed right around the beam pipe and provide high resolution position measurements on tracks close to the interaction point. Silicon detectors are semi-conductor detectors which are modified by doping. For example, doping with Antimony gives an  $n$ -type semi-conductor or with Boron which gives a  $p$ -type semi-conductor. This doped silicon is then used to create a  $p-n$  junction, to which a very large reverse-bias voltage is applied. This

creates a "depletion zone" and once the silicon device is fully depleted we are left with an electric field. When charged particles cross the detector they ionize the depletion zone and create an electrical signal. Silicon detectors come in two varieties, either metal strips:



or pixels:



which provide much higher granularity and a higher precision set of measurements. These detectors are radiation hard and are important for detection secondary vertices close to the primary interaction. Silicon is now the dominant sensor material in use for tracking detectors at the LHC.

The entire tracking volume is usually enclosed in a uniform magnetic field  $B$  and from the curvature of tracks one measures the particle's momentum. The momentum and charge of the particle is measured using a few points of the particle's track (trajectory), which we can use to reconstruct the curvature of the track. The transverse momentum ( $p_T$ ) of charged particles is proportional to the radius of curvature and to the  $B$  field. In particular,

$$p_T = 0.3qBr , \quad (16.5)$$

where the reconstructed track  $p_T$  is measured in  $\text{GeV}/c$ ,  $B$  is in Tesla, the total particle charge is  $qe$  ( $e$  is the magnitude of the electron charge) and  $r$  is measured in meters and is the radius of curvature of the track.

Drift chambers are the most versatile of all detector elements. In addition to measuring momentum and charge, tracks left by charged particles can be extrapolated back to the interaction point. Tracks with significant impact parameters to the beam crossing point, or that can be combined to form a displaced vertex.

### 16.2.2 Electromagnetic Calorimeters

Most experiments have some form of electromagnetic calorimeter. Electromagnetic calorimeters are designed to measure the energy of electromagnetic particles (both *charged* and *neutral*) and their position. This is done by constructing them using a heavy, high  $Z$  material to initiate an electromagnetic shower to totally absorb the energy and stop the particles. The important parameter for the material used in electromagnetic calorimeters is the radiation length  $X_0$ , and have typical values of  $15 - 30 X_0$ . Additionally, it is key to have as little material before the calorimeter as possible (this means the tracker) so that the particles do not radiate before they reach it.

The relative energy uncertainty (or resolution),  $\sigma_E$ , of calorimeters decreases with the energy  $E$  of the particle and can be parameterized as follows:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}, \quad (16.6)$$

where  $a$  is referred to as the stochastic term and quantifies statistics related fluctuations,  $b$  is the constant term and  $c$  is primarily due to noise (for example, in the electronics). The three terms in this equation are added in quadrature (denoted by the symbol  $\oplus$ ).

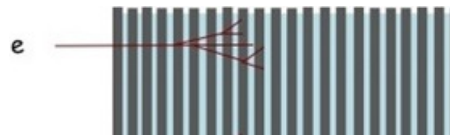
There are two types of calorimeter detectors:

- *Homogeneous Calorimeters* are generally made of an inorganic heavy, high  $Z$  material which is also scintillating. The idea is to create an entire volume to generate the electromagnetic signal, as seen in the figure below:



Examples of these calorimeters include a variety of crystals such as  $CsI$ ,  $NaI$ , and  $PbWO$ , and ionizing noble liquids such as liquid  $Ar$ . The entire photons shower is contained in the uniform crystal blocks with dimensions typically  $5\text{ cm} \times 5\text{ cm} \times 30\text{ cm}$  deep. Adjacent blocks are summed to reconstruct the shower. Energy resolutions of these types of detectors are typically  $\sigma_E/E \sim 1\%$ .

- *Sampling Calorimeters* are made of an active medium which generates signal and a passive medium which functions as an absorber (typically a lead-scintillator sandwich) as seen in the figure:



Examples of active medium materials are scintillators, ionizing noble liquids, and a Cherenkov radiator. The passive material is one of high density, such as lead, iron, copper, or depleted uranium. Energy resolutions of sampling calorimeter detectors are typically  $\sigma_E/E \sim 10\%$ .

The scintillating light created in calorimeters is interpreted as a signal using photo-multiplier tubes (PMT's) and translated as the energy of the particle.

Electromagnetic calorimeters are also very powerful as  $e^-$  detectors. An electron is identified by matching the energy of a shower in the calorimeter to the momentum measured on a charged track pointing to the cluster. Electrons are easily separated from hadrons and muons, which deposit much less energy.

### 16.2.3 Hadron Calorimeters

The purpose of hadronic calorimeters is to measure the energy of heavy hadronic particles. They are similar to electromagnetic calorimeters but bigger. Hadron calorimeters typically are sampling calorimeters and tend to be



larger and coarser in sampling depth than electromagnetic calorimeters, and therefore have larger energy resolutions. For example, the stochastic term is usually in the 30 – 50% or even higher. When a strongly interacting particle goes through material, there are elastic and inelastic interactions with nuclei in the material, producing secondary hadrons. Hadron calorimeters typically use a sampling technique with plates of a dense high  $Z$  material, such as uranium or iron sandwiching a scintillating material, or ionization detector where the shower is sampled. Again, the idea is to get a particle to give up all its energy in the calorimeter. Typical hadronic interaction lengths of materials, such as iron are 15 – 20 cm, and many interaction lengths are needed for an efficient detector. Since the hadron calorimeter in a colliding beam detector has to go outside the drift chamber and electromagnetic calorimeter, this can be a lot of iron!

The most important use of a hadron calorimeter is to measure the energy of dense jets of particles.

An important use of both electromagnetic and hadron calorimeters is to detect neutrinos in an event. Neutrinos will leave no measurable signal in the detector, so the only hope is to detect them indirectly. The experiments use a missing momentum technique. I mentioned that at a hadron machine, the CM of the parton-parton collision is not necessarily the Lab frame. Since fragments of the parent proton and antiproton escape down the beam pipe in the very forward direction, there is no way to use conservation of total momentum in the event to infer the momentum of the unobserved neutrino. However, the components of the momentum in the plane transverse to the beam line ( $p_T$ ) can be measured for all the observed decay products by using the vector sum over the energy deposited in the calorimeters, and that should be zero before and after the collision. Therefore, the neutrino transverse momentum can be inferred as the negative of the vector sum of all the transverse momenta detected in the event.

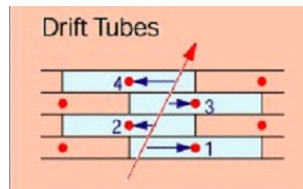
#### 16.2.4 Muon Detectors

Muons are extraordinarily penetrating and therefore the detectors for identifying them are the outer-most layer of a collider detector. Any charged particle that makes it through that many interaction lengths of material is identified as a muon. These detectors are made up of several layers of tracking chambers. Their primary purpose is to measure the momentum and

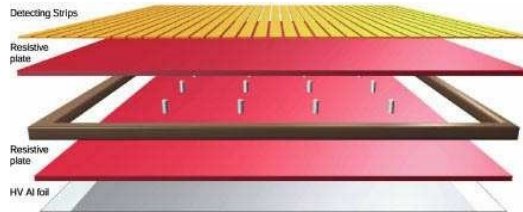
charge of muons. The measurements from the muon chambers are combined with the tracks reconstructed with the inner tracker to fully reconstruct the muon trajectory.

Muon chambers in LHC experiments are made from a series of different types of tracking chambers for precise measurements and some examples include:

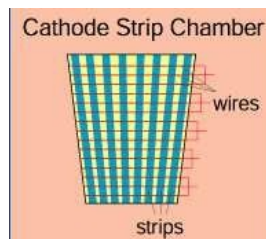
- *Drift Tubes* (DT's): Wire chamber devices, so when muons traveling through kick off atomic electrons in the gas and drift to the positively charged wire.



- *Cathode Strip Chambers* (CSC's): Wires crossed with metallic strips in a gas volume, so when muons traverse the detectors electrons drift to the positively charged wire as described above. Additionally, the positive ions in the gas drift to the metallic strips and induce a charged pulse perpendicular to the wire, giving a two dimensional coordinate of the traveling muon.



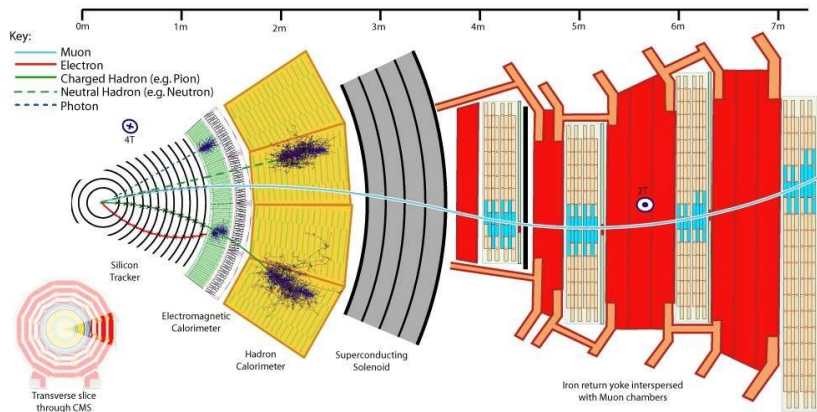
- *Resistive Plate Chambers* (RPC's): Oppositely charged parallel plates containing a gas volume, creating drift electrons when muons cross the detector.



### 16.2.5 Particle Identification

Many experiments find it useful to distinguish protons, pions and kaons. There are currently experiments with very sophisticated particle identification systems based on differences in the pattern of Cerenkov radiation emitted by the various particle species. Low energy experiments can get some information from time of flight or ionization losses in their drift chambers, but the information is limited.

Figure below shows a schematic of a transverse slice of the CMS detector at CERN outlining the identification of various particles.



- *Electrons and Photons:* Electrons are identified as an energy deposit in the electromagnetic calorimeter, and is required to have a shower shape (energy loss) consistent with an electromagnetic shower. It is also required to have little or no energy in the hadronic calorimeter. Since electrons are charged particles it needs to be associated with a track reconstructed in the tracker, and is therefore required to have a matched position measurement in the calorimeter with the one from the track. If the electromagnetic cluster of energy does not have a track pointing to it, then it becomes a candidate for being a photon.
- *Muons:* Muon identification begins by reconstructing a track in the muon system which is then matched with a track in the inner tracker. Additionally, since muons are minimum ionizing particles, they are expected to deposit little or no energy in the calorimeters.

- *Jets*: Jets are created when a quark or gluon is knocked out of the proton and due to parton confinement subsequently a hadron is created. This hadron forms a jet once it decays and fragments into many particles (hadronization), which are essentially collimated object. The reconstruction of a jet is the experimentalists representation of a parton. There are several algorithms for reconstructing jets but overall what these reconstruction algorithms do is attempt to group the particles from the hadronization process together and measure the energy of the parton. There are two main categories of jet algorithms that experimentalists and theorist use to reconstruct jets: (i) Cone algorithms, when one draws circles of  $\Delta R$  around clusters of energy according to some rule, and (ii) Recursive luster reconstruction such as the anti- $k_T$  algorithm, which is now the default jet algorithm of choice for the LHC experiments.

Measuring the jet energy has several challenges since it is impossible to determine which particles came from which hadronization process. There are several effects which contribute to the complication of the jet energy measurement, such as multiple  $pp$  interactions, spectator partons interacting and noise in the calorimeters. However, experimentalists have ways of correcting for such effects and this calibration the jet energy is generally called the *Jet Energy Scale* (JES) and often depends on the  $p_T$  and the  $\eta$  of the jet.

- *b-Hadrons*: There is a special category of jets coming from  $b$  hadrons which are long-lived ( $\sim 450 \mu m$ ) and massive. There are two standard techniques for identifying a  $b$  hadron decay, referred to as *b-tagging*. One can look for displaced tracks forming a secondary vertex away from the primary vertex of the interaction. Alternatively one can identify soft leptons (electrons or muons) inside the jet, which would be a signature specific to semi-leptonic  $b$  decays.
- *Tau Leptons*: The identification of tau leptons is for hadronically decaying taus, which decay  $\sim 49\%$  of the time to a single charged hadron and neutrinos and  $\sim 15\%$  of the time to three charged hadrons and neutrinos. Leptonically decaying taus are indistinguishable from "normal" electrons and muons. The reconstruction algorithms for taus assume that taus form narrow jets in the calorimeter. First one forms a  $\Delta R$  cone around clusters of energy and tracks (a signal cone) and a second larger  $\Delta R$  one around the signal cone (an isolation cone) where there is little or no calorimeter and track activity. In the signal cone,

one or three tracks are required as well as electromagnetic energy in the calorimeters from neutral particles (such as  $\pi^0$ s).

- *Neutrinos*: Neutrinos are weakly interacting particles and pass through all the material in the LHC detectors. They are identified indirectly by the imbalance of energy in calorimeters. This missing energy is one of the most interesting and most difficult quantities for experimentalists. Various effects could contribute to the complications of the measurement such as dead calorimeter cells or a jet whose hardest hadron enters a crack (between cells) in the calorimeter or an improperly calibrated calorimeter. Therefore, we need to carefully understand this quantity as it is very important for searches of new physics processes which would produce additional weakly interacting particles.

### 16.2.6 Selection of Events and Triggers

The first thing that the experiment has to decide is whether or not an interaction of interest has occurred at a particular beam crossing, or beam spill. This decision is crucial. If something interesting happens, then the event will be read out, which takes time (meaning subsequent events will be missed). In the trade, the process by which the experiment decides whether or not an event is interesting is called *the trigger*. Too loose or indiscriminate of a trigger will result in lots of dead time for the experiment, so good data will be lost. Too tight or selective of a trigger means interesting physics may be thrown away.

For  $e^+e^-$  machines, triggering for most types of events is quite straight forward. Cross sections are low. Fairly simple requirements requiring evidence of a minimum number of charged tracks in the detector, or a minimum threshold for energy deposited in the calorimeter, will yield a trigger that is essentially without dead time, but still preserves 99% efficiency for  $e^+e^-$  annihilation events.

For  $pp$  machines and fixed target experiments, the trigger is difficult and must be carefully thought through. Kinematics helps because heavy objects will not have significant boost in the Lab frame. When a heavy object is produced, its decay products can have a lot of energy transverse to the beam direction, while the uninteresting events send most of the beam energy down the beam pipe. It is only by a careful selection process that these events can

be identified. However, it is an amusing exercise to speculate what is going on in individual events. Looking at event pictures is fun, instructive, and keeps us attached to the real world, but it is not how we do physics!

At design a CM energy of 14 TeV and a luminosity of  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$ , the total cross section at the LHC is  $\sim 10^8 \text{ nb}$ . The rate for all collisions will be around 40 MHz. Since it is not possible to record every collision event, quick decisions need to be made a priori selecting the interesting events worthy of analysis. This filter, or *trigger*, needs to single out rare processes and reduce the common processes. We also want to keep less interesting events for "standard-candle" measurements (such as jet and  $W$  boson and  $Z$  boson production cross sections), calibrations, and so on. It is critical to consider carefully the make-up of the trigger and make wise choices, otherwise the events will be thrown away forever.

A typical trigger table will contain triggers on: electroweak particles (photons, electrons, muons, taus) at as low an energy as possible, very high-energy partons (jets), and apparent invisible particles.

Theory very often plays a role in guiding these choices, therefore it is important to have good communication between theorists and experimentalists.

The LHC experiments have two levels of triggers:

- L1, which bases its decision on hardware electronics;
- L2, which based on software programming (the high level trigger, or HLT).

Recall, the starting trigger rate is 40 MHz, which gets reduces after the L1 trigger to a rate of around 100 kHz. The HLT trigger further prunes this down to roughly 150 – 200 Hz, which is the event rate that the experiments record. Therefore, the final decision of the trigger is to keep  $\sim 1/200,000$  events occur every second, there is no room for mistakes.

One should be very aware that all measurements are distorted by the trigger selection thresholds and any measurement must account for the efficiency of the trigger and that resulting distortion. Therefore, it is necessary to include "backup" or "monitoring" trigger for measuring the efficiencies of the more interesting triggers to be used for physics analysis.

## 16.3 Data Analysis

A typical experiment may have millions of events recorded, but physics analysis may end up with a few hundred events. An analysis searching for a rare process may end up with a sample of only 10 or 20 events. One has to develop a procedure to select events characteristic of the physics process one wants to study but without unnecessary bias. This is an extraordinary challenge, especially when one considers the magnitude of the winnowing that must occur.

The LHC produces roughly 15 petabytes (15 million gigabytes) of data annually. Finally, there is the challenging task of distributing the recorded data around the world for analysis. The LHC has a tiered computing model to distribute this data around the world, referred to as *the Grid*.

### 16.3.1 Monte Carlo

The primary tool that experimenters have to help them develop a selection procedure is called *the Monte Carlo* (MC). The MC has two parts: *the physics simulation* and *the detector simulation*.

A ubiquitous problem in particle physics is the following: given a source located some distance from a detector, predict the number of counts that should be observed within the solid angle spanned by the detector (or within a bin of its phase-space acceptance), as a function of the properties of the source, the intervening medium, and the efficiency of the detector. Essentially, the task is to compute integrals of the form

$$N_{\text{Count}}(\Delta\Omega) = \int_{\Delta\Omega} d\Omega \frac{d\sigma}{d\Omega}, \quad (16.7)$$

with  $d\sigma$  a differential cross section for the process of interest.

In particle physics, phase space has three dimensions per final-state particle (minus four for overall 4-momentum conservation). Thus, for problems with more than a few outgoing particles, the dimensionality of phase space increases rapidly.

The task of a *MC Event Generator* is to simulate and calculate everything that happens in a high-energy collision. Starting from a differential cross

Relative uncertainty with $n$ points	1-Dim	$d$ -Dim	$n_{eval}/point$
Trapezoidal Rule	$\frac{1}{n^2}$	$\frac{1}{n^{2/d}}$	$2^d$
Simpson's Rule	$\frac{1}{n^4}$	$\frac{1}{n^{4/d}}$	$3^d$
Monte Carlo	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	1

Table 16.1: *Relative uncertainty after  $n$  evaluations, in 1 and  $d$  dimensions, for two traditional numerical integration methods and stochastic MC. The last column shows the number of function evaluations that are required per point, in  $d$  dimensions.*

section that describes our best understanding of the physics happening at a given energy. An event generator will generate 4-momentums for a properly distributed sample of events according to whatever model (such as the SM) we specified. The SM tells us how to distribute the 4-vectors of initial particles. We then need some model of intermediate processes according to whatever we know about their branching ratios and lifetimes. This procedure keeps going until one has a set of 4-vectors for long-lived particles that will actually end up in the detector.

*A word of warning:* obviously, the physics given by MC requires some compromises to be made, it is only as good as the physics we put into it! If we have neglected some physics in the MC that is present in the data, we will get discrepancies between what the MC thinks we should be seeing in our data, and the data we collect. It is always important to understand what the limitations in the physics inputs to simulations are.

The standard  $1d$  numerical-integration methods give very slow convergence rates for higher-dimensional problems. For illustration, a table of convergence rates in 1 and  $d$  dimensions is given in Tab. 16.1, comparing the Trapezoidal (2-point) rule and Simpson's (3-point) rule to random-number-based MC. In  $1d$ , the  $1/n^2$  convergence rate of the Trapezoidal rule is much faster than the stochastic  $1/\sqrt{n}$  of random-number MC, and Simpson's rule converges even faster. However, as we go to  $d$  dimensions, the convergence





"This risk, that convergence is only given with a certain probability, is inherent in Monte Carlo calculations and is the reason why this technique was named after the world's most famous gambling casino. Indeed, the name is doubly appropriate because the style of gambling in the Monte Carlo casino, not to be confused with the noisy and tasteless gambling houses of Las Vegas and Reno, is serious and sophisticated."

Figure 16.1: *Left*: The casino in Monaco. *Right*: Extracted from F. James, "Monte Carlo theory and practice", Rept. Prog. Phys. **43** (1980) 1145.

rate of the  $n$ -point rules all degrade with  $d$  (while the number of function evaluations required for each "point" simultaneously increases). The MC convergence rate, on the other hand, remains the simple stochastic  $1/\sqrt{n}$ , independent of  $d$ , and each point still only requires one function evaluation.

These are the main reasons that MC is the preferred numerical integration technique for high-dimensional problems. In addition, the random phase-space vectors it generates can be re-used in many ways, for instance as input to iterative solutions, to compute many different observables simultaneously, and/or to hand "events" to propagation and detector-simulation codes.

Therefore, virtually all numerical cross section calculations are based on MC techniques in one form or another, the simplest being the RAMBO algorithm, which can be expressed in about half a page of code and generates a flat scan over  $n$ -body phase space. Strictly speaking, RAMBO is only truly uniform for massless particles. Its massive variant makes up for phase-space biases by returning weighted momentum configurations.

In some cases, e.g. in QCD, simple algorithms like RAMBO become inefficient and one needs to introduce generic (i.e. automated) importance-sampling methods, such as offered by the VEGAS algorithm. This is still the dominant basic technique, although most modern codes do employ several additional refinements, such as several different copies of VEGAS running in parallel (multi-channel integration), to further optimise the sampling. Alternatively, a few algorithms incorporate the singularity structure explicitly in their phase-space sampling, called SARGE, or by using all-orders Markovian parton showers to generate them (VINCIA).

The price of using random numbers is that we must generalise our notion of convergence. In calculus, we say that a sequence  $\{A\}$  *converges* to  $B$  if an  $n$  exists for which the difference  $|A_{i>n} - B| < \epsilon$  for any  $\epsilon > 0$ . In random-number-based techniques, we cannot completely rule out the possibility of very pathological sequences of "dice rolls" leading to large deviations from the true goal, hence we are restricted to say that  $\{A\}$  converges to  $B$  if an  $n$  exists for which *the probability* for  $|A_{i>n} - B| < \epsilon$ , for any  $\epsilon > 0$ , is greater than  $P$ , for any  $P \in [0, 1]$ . This risk, that convergence is only given with a certain probability, is the reason why MC techniques were named after the famous casino in Monaco, illustrated in Fig. 16.1.

The second part of the Monte Carlo is to simulate how events will appear in the detector (*the detector simulation*). Here one takes the 4-vectors of the stable (long lived) particles produced by the physics simulation and propagates them through the detector. The detector simulation, for example, simulates the multiple Coulomb scattering and energy losses as the particle passes through the beam pipe. It propagates the particle through the drift chamber and generates MC data as simulated signals on drift chamber sense wires. It simulates the electromagnetic showers in the calorimeter and so on. The detector simulation is hugely expensive in terms of computer time.

The great value of MC is that one can generate a sample of "fake data" or "MC data" to test an analysis procedure on. One can determine the effect that analysis selection criteria will have on efficiency, one can study potential backgrounds. The most important function of MC is that one can, in an unbiased way, come up with criteria to select a signal. It is easy when one is looking for rare processes with small numbers of signal events and with large backgrounds to end up enhancing a statistical fluctuation. The only way to avoid that is to use MC data to determine event selection and background suppression techniques before ever looking at the data.

### 16.3.2 What Can Go Wrong

Experimental result can be made to look much better with different binning. That is considered cheating if it affects the signal yield extracted by the analysis. The yields should be tested that they were independent of bin size. The evaluation of the background can be done in many different ways (from data and MC) and one needs to check the result are stable when the cuts are changed. These are all checks you should expect to see experimenters

do. So what are the questions you should ask when deciding whether to believe a marginal ( $3-4\sigma$ ) result or in deciding whether you believe the level of precision on a more significant result:

- How were event selection criteria determined?
- How was the background evaluated?
- What happens when the event selection cuts are varied?
- What is the error on the efficiency and how was it determined?

There are some other pitfalls you should be aware of:

1. Sometimes experimenters do not understand their detectors as well as they think they do.
2. Physicist are prone to over-averaging, where many measurements are averaged, weighted by their combined statistical and systematic error. The Particle Data Group contributes to this problem in some ways by making the data so easily available.

A famous example of over averaging was the  $\tau$  1-prong problem. The  $\tau$  lepton can decay just like the muon ( $\tau \rightarrow e\nu\bar{\nu}$ ,  $\tau \rightarrow \mu\nu\bar{\nu}$ ), but it can also decay to hadrons. In 1984, it was noticed that if one took the world averages for the measured  $\tau$  branching ratios and used theoretical predictions to constrain poorly measured modes, then the measured inclusive 1-prong branching ratio was significantly larger than the sum of the exclusive modes. In 1992 one saw that, taking world averages, one found a significant discrepancy in the inclusive rate and the sum of the exclusive rates.

For years there were speculations about new physics and unseen decay modes. The problem, however, was caused by averaging many experiments with large errors and extracting an average with rather small errors. It is very dangerous to take results from 10 different measurements with roughly equal precision and then average them to get a factor of three smaller error. If errors were statistical only, there would not be a difficulty. The problem comes from *systematic errors* which may be correlated experiment to experiment. Systematic errors are hard to evaluate and correlations are hard to spot. For example, there

may be unknown correlated errors due to incorrect inputs or overlooked backgrounds. So global averages need to be done with great care.

3. A third pitfall that affects experimenters more than theorists is the enormous temptation to stop at the "right answer". We very often have a preconceived prejudice on what a result should be. A good experimenter does many of the systematic studies and checks before looking at the actual number he or she is getting.
4. Finally, theorists tend to fall into the "single event" pit. What does it mean to find a single event? In 1964, the  $\Omega^-$  was discovered with the observation of one event. However, there was an enormous amount of information in the bubble chamber photograph that captured that one event. The decay was fully reconstructed  $K^-p \rightarrow \Omega^- K^+ K^0$ ,  $\Omega^- \rightarrow \Xi^0 \pi^-$ ,  $\Xi^0 \rightarrow \Lambda^0 \pi^0$ ,  $\Lambda^0 \rightarrow \pi^- p$ ,  $\pi^0 \rightarrow \gamma_1 \gamma_2$ ,  $\gamma_1 \rightarrow e^+ e^-$ ,  $\gamma_2 \rightarrow e^+ e^-$ , except for the  $K^0$ . They were able to claim discovery because there was so much information in the event that the probability for the background to produce such an event was vanishingly small. However, for modern collider experiments, it is impossible to have the same level of information. The crucial issue is not how many events one finds, but how well the background can be evaluated and understood, and what is the probability that a background process could imitate the event one is looking for.

**Exercise 16.1:** Imagine a hadron collider such as the LHC runs for one year with an instantaneous luminosity of  $10^{31} \text{cm}^{-2} \text{s}^{-1}$ , how much integrated luminosity will be delivered to an experiment?

**Exercise 16.2:** In  $100 \text{pb}^{-1}$  of data, how many  $p\bar{p} \rightarrow t\bar{t}$  events will be produced at the LHC at  $\sqrt{s} = 7 \text{TeV}$ ?

**Exercise 16.3:** What size beam spot is needed for the luminosity  $L = 1 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$  at the LHC?

**Exercise 16.4:** How would you detect  $500 \text{MeV}$   $\gamma$ -rays? With: (a) Hydrogen bubble chamber; (b) Shower counter; (c) Geiger counter.

**Exercise 16.5:** The energy loss of an energetic muon in matter is due mainly to collisions with: (a) nucleons; (b) nuclei; (c) electrons.

**Exercise 16.6:** How are the pions, muons and electrons distinguished in photographic emulsions and in bubble chambers?

## Chapter 17

# Limitations of Particle Accelerators

### 17.1 Limit I: The Geography

Beyond physical and technical limits there is a serious boundary condition for accelerators – the landscape. For a given technology, pushing the particle energy of a storage ring to higher values will necessitate larger machines and we may suddenly encounter the problem that our device no longer fits even in the entire region surrounding our facility. As regards LHC, the largest storage ring at present, the space between Lake Geneva and the Jura mountains defined the size of the tunnel for the present LHC. As a consequence, the maximum feasible beam energy available for high-energy physics experiments is determined by the geographical boundary conditions.

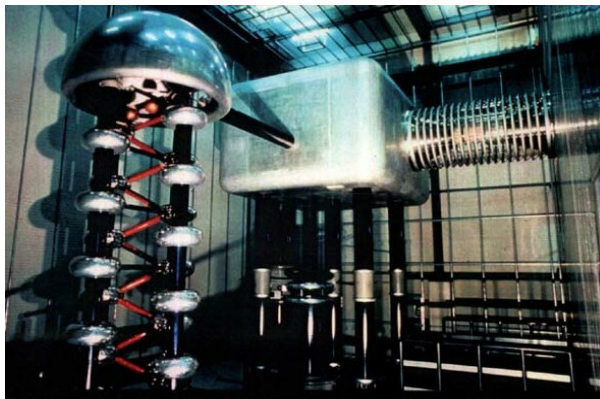


Certainly, there are extremely high particle energies in cosmic rays, but you will agree that the accelerators driving these are also much larger!

## 17.2 Limit II: Voltage Breakdown in DC Accelerators

Let us take a brief look at the path paved by the ingenious scientific developments dating back to the discovery of the nucleus by Rutherford. Using alpha particles on the level of MeV is not ideal for precise, triggerable and healthy experiments. So Rutherford discussed with two colleagues, Cockcroft and Walton, the possibility of using artificially accelerated particles. Based on this idea, within only four years Cockcroft and Walton invented the first particle accelerator ever built, and in 1932 they gave the first demonstration of the splitting of a nucleus (lithium) by using a 400 keV proton beam.

Cockcroft and Walton were awarded the Nobel Prize for their invention. Their acceleration mechanism was based on a rectifier or Greinacher circuit, consisting of a number of diodes and capacitors that transformed a relatively small AC voltage to a DC potential which corresponds, depending on the number of diode/capacitor units used, to a multiple of the applied basic potential. The particle source was a standard hydrogen discharge source connected to the high-voltage part of the system, and the particle beam was accelerated to ground potential, hitting the lithium target. A photograph of a Cockcroft-Walton generator that was used at CERN for many years as a pre-accelerator for the proton beams (the device has since been replaced by the more compact and efficient RFQ technique) is presented below.



### 17.3. LIMIT III: SIZE OF THE AC ACCELERATING STRUCTURES<sup>261</sup>

In parallel to Cockcroft and Walton, but based on a completely different technique, another type of DC accelerator had been invented: Van de Graaff designed a DC accelerator that used a mechanical transport system to carry charges, sprayed on a belt or chain, to a high-voltage terminal.

In general these machines can reach higher voltages than the Cockcroft-Walton devices, but they are more limited in terms of particle intensity. Common to all DC accelerators is the limitation on the achievable beam energy due to high-voltage breakdown effects (discharges). Without using an insulating gas ( $SF_6$  in most cases), electric fields will be limited to about  $1 \text{ MV m}^{-1}$ , and even with the most sophisticated devices, acceleration voltages on the order of MV cannot be overcome. In fact, the example of a Van de Graaff accelerator below.

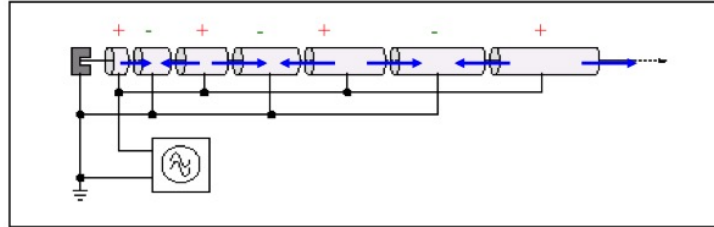


This is the reliable machine for precise measurements in atomic and nuclear physics, shows an approach that has been applied in a number of situations: injecting a negative ion beam (even  $H^-$  is used) and stripping the ions in the middle of the high-voltage terminal allows one to profit from the potential difference twice and thus to make another step of gain in beam energy.

### 17.3 Limit III: Size of the AC Accelerating Structures

Given the obvious limitations of the DC machines described above, the next step forward is natural. In 1928, Widerøe developed the concept of a AC accelerator. Instead of rectifying the AC voltage, he connected a series of

acceleration electrodes in an alternating manner to the output of an AC supply. The schematic layout is shown below,

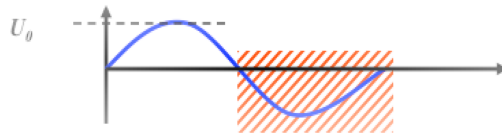


where the direction of the electric field is indicated. In principle, this device can produce step by step a multiple of the acceleration voltage, as long as for the negative half-wave of the AC voltage the particles are shielded from the decelerating field. The energy gain after the  $n$ -th step is therefore

$$E_n = n \cdot q \cdot U_0 \cdot \sin \psi_s , \quad (17.1)$$

where  $n$  denotes the acceleration step,  $q$  the charge of the particle,  $U_0$  the applied voltage per gap, and  $\psi_s$  the phase between the particle and the changing AC voltage.

A key quantity in such a Widerøe structure is the length of the drift tubes that will protect the particles from the negative half-wave of the sinusoidal AC voltage. For a given frequency of the applied radio-frequency (RF) voltage, the length of the drift tube is defined by the speed of the particle and the duration of the negative half-wave of the sinusoidal voltage.



The time-span of the negative half-wave is defined by the applied frequency,  $\Delta t = \tau_{\text{rf}}/2$ , so for the length of the  $n$ -th drift tube we get

$$l_n = v_n \cdot \frac{\tau_{\text{rf}}}{2} . \quad (17.2)$$

Given the kinetic energy of the particle,

$$E_{\text{kin}} = \frac{1}{2} m v^2 , \quad (17.3)$$



we obtain directly that

$$l_n = \frac{1}{\nu_{\text{rf}}} \cdot \sqrt{\frac{nqU_0 \sin \psi_s}{2m}}, \quad (17.4)$$

which defines the design concept of the machine. On the photo of Unilac at the Institute for Heavy Ion Research (Darmstadt, Germany) clearly visible are the structure of the drift tubes and their increasing length as a function of the particle energy.



Two remarks should be made in this context.

- The short derivation here is based on the classical approach, and in fact these accelerators are usually optimum for ‘low-energetic’ proton or heavy-ion beams. Typical beam energies (referring to protons) are on the order of 10 MeV; for example, the present Linac 2 at CERN delivers the protons for LHC operation with an energy of 50 MeV, corresponding to a relativistic  $\beta$  of 0.31.
- For higher energies, even in the case of protons or ions, the speed will at some point approach the speed of light, and the length of the drift tubes and hence the dimension of the whole accelerator will reach a size that may no longer be feasible. More advanced ideas are needed in order to keep the machine within reasonable dimensions, and the next natural step in the historical development was to introduce magnetic fields and bend the particle beam into a circle.

## 17.4 Limit IV: Magnetic Guide Field

A significant step forward in achieving high beam energies involves the use of circular structures. In order to apply over and over again the accelerating fields, we try to bend the particles onto a circular path and so bring them back to the RF structure where they will receive the next step-up in energy. To do this, we introduce magnetic (or electric) fields that will deflect the particles and keep them on a well-defined orbit during the complete acceleration process. The Lorentz force that acts on a particle will therefore have to compensate exactly the centrifugal force due to the bent orbit. In general, we can write

$$\mathbf{F} = q \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) . \quad (17.5)$$

For high-energy particle beams, the velocity  $\mathbf{v}$  is close to the speed of light and so represents a nice amplification factor whenever we apply a magnetic field. As a consequence, it is much more convenient to use magnetic fields for bending and focusing the particles.

Therefore, neglecting electric fields for the moment, we write the Lorentz force and the centrifugal force of the particle on its circular path as

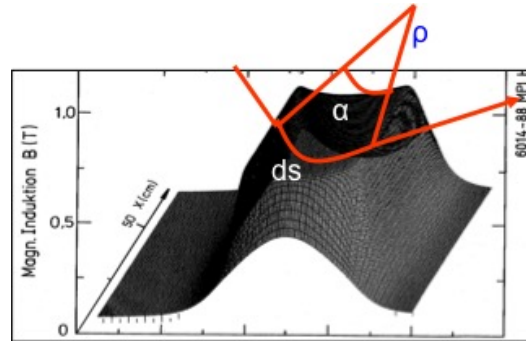
$$\begin{aligned} F_{\text{Lorentz}} &= e \cdot v \cdot B , \\ F_{\text{Centrifugal}} &= \frac{\gamma m_0 v^2}{\rho} , \end{aligned} \quad (17.6)$$

where  $\rho$  is the radius of the particles orbit. Assuming an idealized homogeneous dipole oriented along the particle orbit, we define the condition for a perfect circular orbit as equality between these two forces; this yields the following condition for the idealized ring:

$$\frac{p}{e} = B \cdot \rho , \quad (17.7)$$

where we refer to protons and have accordingly set  $q = e$ . This condition relates the so-called beam rigidity  $B\rho$  to the particle momentum that can be carried in the storage ring, and it will ultimately define, for a given magnetic field of the dipoles, the size of the storage ring.

In reality, instead of a continuous dipole field the storage ring will be built out of several dipoles, powered in series to define the geometry of the ring. For a single magnet, the particle trajectory is shown schematically below.



In the free space outside the dipole magnet, the particle trajectory follows a straight line. As soon as the particle enters the magnet, it is bent onto a circular path until it leaves the magnet at the other side.

The overall effect of the main bending (or "dipole") magnets in the ring is to define a more or less circular path, which we will call the "design orbit". By definition, this design orbit has to be a closed loop, and so the main dipole magnets in the ring have to define a bending angle of exactly  $2\pi$  overall. If  $\alpha$  denotes the bending angle of a single magnet, then

$$\alpha = \frac{ds}{\rho} = \frac{B ds}{B \cdot \rho} . \quad (17.8)$$

We therefore require that

$$\frac{\int B dl}{B \cdot \rho} = 2\pi . \quad (17.9)$$

Thus, a storage ring is not a "ring" in the true sense of the word but more a polygon, where "poly" means the discrete number of dipole magnets installed in the "ring".

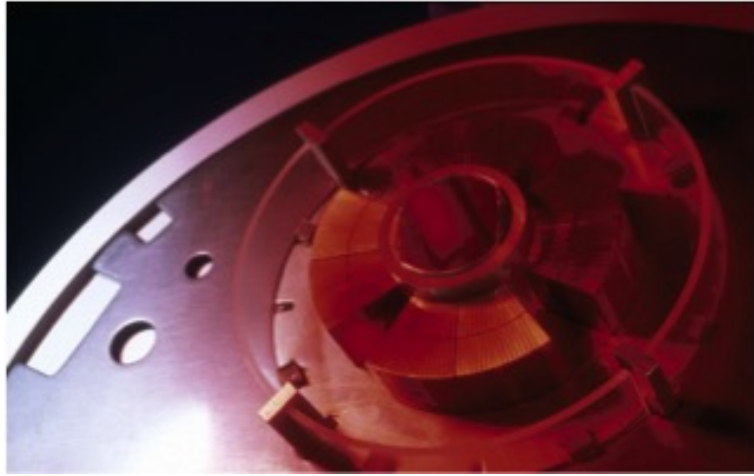
In the case of the LHC, the dipole field has been pushed to the highest achievable values; 1232 superconducting dipole magnets, each of length 15 m, define the geometry of the ring and, via (17.9), the maximum momentum for the stored proton beam. Using the equation given above, for a maximum momentum of  $p = 7$  TeV we obtain a required magnetic field of

$$B = \frac{2\pi \cdot 7000 \cdot 10^9 \text{ eV}}{1232 \cdot 15 \text{ m} \cdot 2.99792 \cdot 10^8 \text{ m s}^{-1}} , \quad (17.10)$$

or

$$B = 8.33 \text{ T} , \quad (17.11)$$

to bend the beams. The photo below shows the LHC dipole magnets, built out of superconducting  $NbTi$  filaments, which are operated at a temperature of  $T = 1.9$  K.



In addition to the main bending magnets that guide the beam onto a closed orbit, focusing fields are needed to keep the particles close together. In modern storage rings and light sources, the particles are kept in the machine for many hours, and a carefully designed focusing structure is needed to maintain the necessary beam size at different locations in the ring.

Following classical mechanics, linear restoring forces are needed, just as in the case of a harmonic pendulum. Quadrupole magnets provide the corresponding property: they create a magnetic field that depends linearly on the particle amplitude, i.e. the distance of the particle from the design orbit:

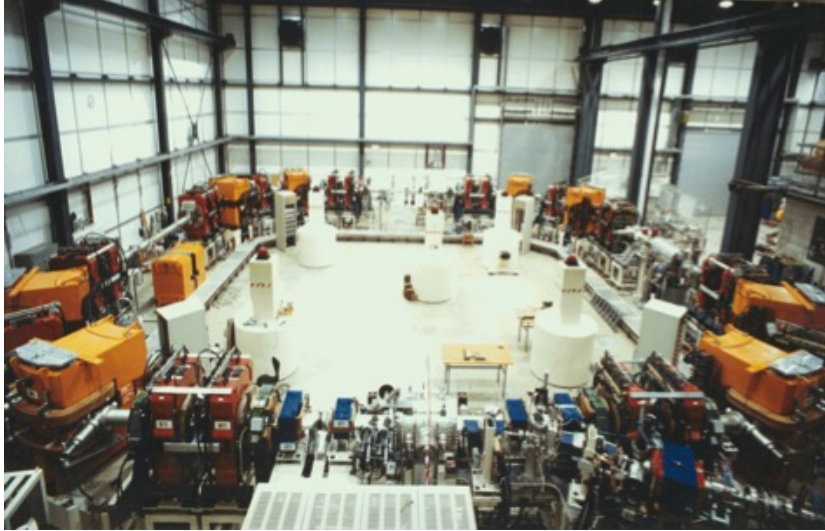
$$B_x = g \cdot y \quad , \quad B_y = g \cdot x \quad . \quad (17.12)$$

The constant  $g$  is called the gradient of the magnetic field and characterizes the focusing strength of the magnetic lens in both transverse planes. For convenience it is (like the dipole field) normalized to the particle momentum. The normalized gradient is denoted by  $k$  and defined as:

$$k = \frac{g}{p/e} = \frac{g}{B\rho} \quad . \quad (17.13)$$

Now that we have defined the basic building blocks of a storage ring, we need to arrange them in a so-called magnet lattice and optimize the field strengths

in such a way as to obtain the required beam parameters. An example of what such a magnet lattice looks like is given below.



This photograph shows the dipole (orange) and quadrupole (red) magnets in the TSR storage ring in Heidelberg. Eight dipoles are used to bend the beam in a ‘circle’, and the quadrupole lenses between them provide the focusing to keep the particles within the aperture limits of the vacuum chamber. As the case of the LHC, the quadrupole magnet, as dipole one, is built in superconducting technology.

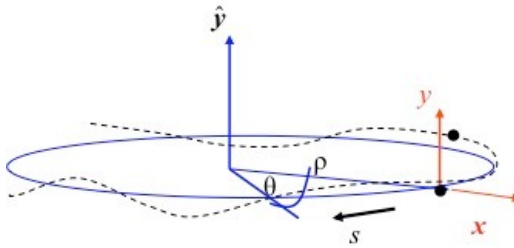
A general design principle of modern synchrotrons or storage rings should be pointed out here. In general, these machines are built following a so-called separate-function scheme: every magnet is designed and optimized for a certain task, such as bending, focusing, chromatic correction, and so on. We separate the magnets in the design according to the job they are supposed to do; only in rare cases a combined-function scheme is chosen, where different magnet properties are combined in one piece of hardware. To express this mathematically, we use the general Taylor expansion of the magnetic field,

$$\frac{B(x)}{p/e} = \frac{1}{\rho} + k \cdot x + \frac{1}{2!}mx^2 + \frac{1}{3!}nx^3 + \dots . \quad (17.14)$$

Following the arguments above, for the moment we take into account only

constant (dipole) or linear (quadrupole) terms. The higher-order field contributions is treated as small perturbations.

The particles will now follow the "circular" path defined by the dipole fields, and in addition will undergo harmonic oscillations in both transverse planes. The situation is shown schematically below.



An ideal particle will follow the design orbit that is represented by the circle in the diagram. Any other particle will perform transverse oscillations under the influence of the external focusing fields, and the amplitude of these oscillations will ultimately define the beam size.

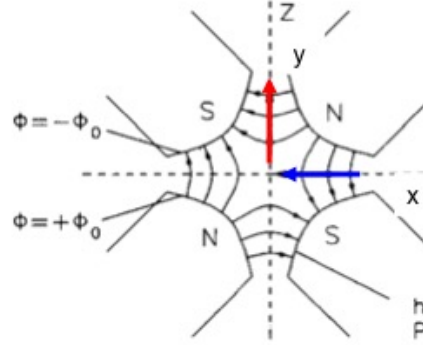
Unlike a classical harmonic oscillator, however, the equations of motion in the horizontal and vertical planes differ somewhat. Assuming a horizontal focusing magnet, the equation of motion is

$$x'' + x \cdot \left( \frac{1}{\rho^2} + k \right) = 0 , \quad (17.15)$$

where  $k$  is the normalized gradient introduced above and the  $1/\rho^2$  term represents the so-called weak focusing, which is a property of the bending magnets. In the vertical plane, on the other hand, due to the orientation of the field lines and by Maxwell's equations, the forces instead have a defocusing effect; also, the weak focusing term disappears:

$$y'' - y \cdot k = 0 . \quad (17.16)$$

The principal problem arising from the different directions of the Lorentz force in the two transverse planes of a quadrupole field is sketched below, showing field configuration in a quadrupole magnet and the direction of the focusing and defocusing forces in the horizontal and vertical planes.



It is the task of the machine designer to find an adequate solution to this problem and to define a magnet pattern that will provide an overall focusing effect in both transverse planes.

Following closely the example of the classical harmonic oscillator, we can write down the solutions of the above equations of motion. For simplicity, we focus on the horizontal plane; a "focusing" magnet is therefore focusing in this horizontal plane and at the same time defocusing in the vertical plane. Starting with initial conditions for the particle amplitude  $x_0$  and angle  $x'_0$  in front of the magnet element, we obtain the following relations for the trajectory inside the magnet:

$$\begin{aligned} x(s) &= x_0 \cdot \cos(\sqrt{|K|} s) + x'_0 \cdot \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|} s) , \\ x'(s) &= -x_0 \cdot \sqrt{|K|} \sin(\sqrt{|K|} s) + x'_0 \cdot \cos(\sqrt{|K|} s) . \end{aligned} \quad (17.17)$$

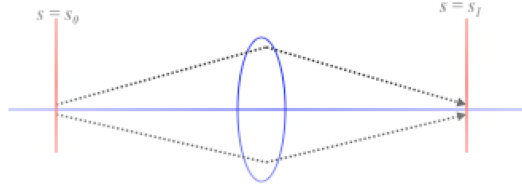
Here the parameter  $K$  combines the quadrupole gradient and the weak focusing effect,  $K = k - 1/\rho^2$ . Usually these two equations are combined into a more elegant and convenient matrix form:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{foc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0 , \quad (17.18)$$

where the matrix  $\mathbf{M}_{\text{foc}}$  contains all the relevant information about the magnet element,

$$\mathbf{M}_{\text{foc}} = \begin{pmatrix} \cos(\sqrt{|K|} s) & \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|} s) \\ -\sqrt{|K|} \sin(\sqrt{|K|} s) & \cos(\sqrt{|K|} s) \end{pmatrix} . \quad (17.19)$$

Schematic principle of the effect of a focusing quadrupole magnet is visualized below.



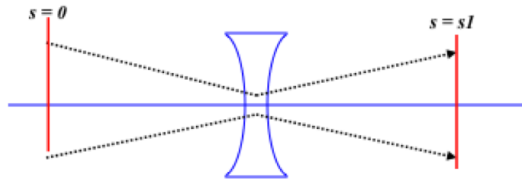
In the case of a defocusing magnet, we obtain analogously that

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{defoc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0, \quad (17.20)$$

with

$$\mathbf{M}_{\text{defoc}} = \begin{pmatrix} \cosh(\sqrt{|K|} s) & \frac{1}{\sqrt{|K|}} \sinh(\sqrt{|K|} s) \\ \sqrt{|K|} \sinh(\sqrt{|K|} s) & \cosh(\sqrt{|K|} s) \end{pmatrix}. \quad (17.21)$$

The schematic principle of the effect of a defocusing quadrupole magnet is:

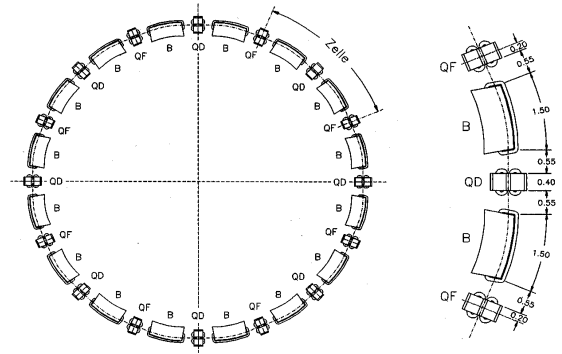


For completeness, we also include the case of a field-free drift, with  $K = 0$ .

$$\mathbf{M}_{\text{drift}} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}. \quad (17.22)$$

This matrix formalism allows us to combine the elements of a storage ring in an elegant way and it is straightforward to calculate the particle trajectories. As an example, we consider the simple case of an alternating focusing and defocusing lattice, a so-called FODO lattice: a simple periodic chain of bending magnets and focusing/defocusing quadrupoles forming the basic structure of a storage ring.

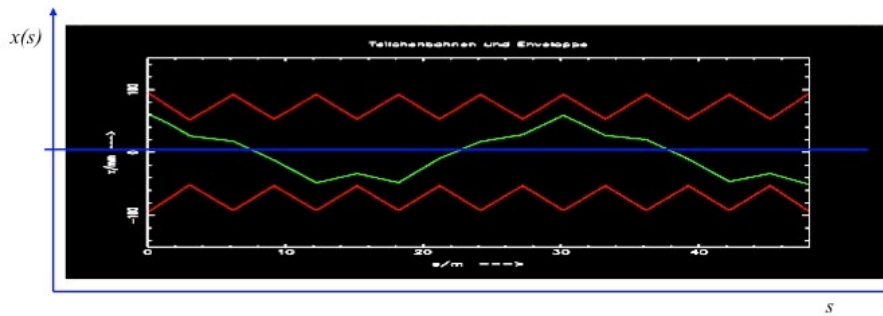




As we know the properties of elements in the accelerator, we can construct the corresponding matrices and calculate step by step the amplitude and angle of a single particle's trajectory around the ring. Even more conveniently, we can multiply out the different matrices and, given initial conditions  $x_0$  and  $x'_0$ , obtain directly the trajectory at any location in the ring:

$$\mathbf{M}_{\text{total}} = \mathbf{M}_{\text{foc}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{dipole}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{defoc}} \cdots . \quad (17.23)$$

The trajectory thus obtained is shown schematically below.

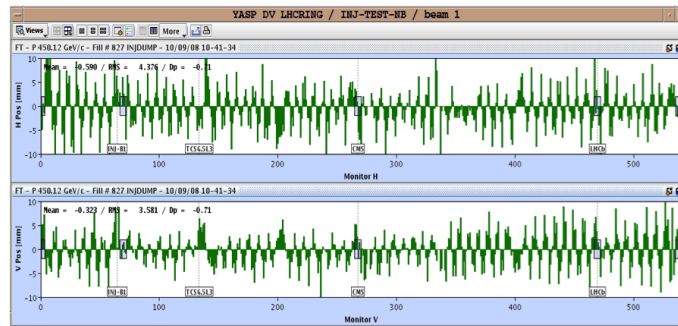


We emphasize the following facts in this context:

- At each moment, or in each lattice element, the trajectory is a part of a harmonic oscillation;
- Due to the different restoring or defocusing forces, the solution will look different at each location;

- In the linear approximation that we make use of in this context, all particles experience the same external fields, and their trajectories will differ only because of their different initial conditions;
- There seems to be an overall oscillation in both transverse planes while the particle is travelling around the ring. Its amplitude stays well within the boundaries set by the vacuum chamber.

Coming closer to a real existing machine, we show the orbit measured during one of the first injections into the LHC storage ring.

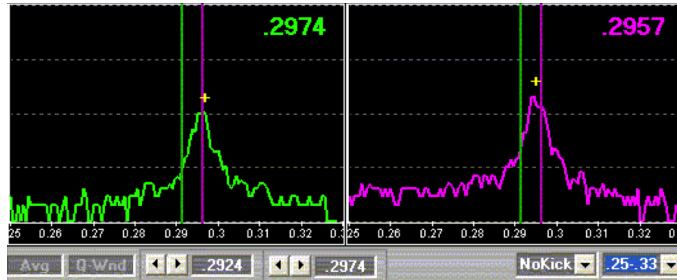


The horizontal oscillations are plotted in the upper half of the figure and the vertical oscillations in the lower half (scale  $\pm 10$  mm). Each histogram bar indicates the value recorded by a beam position monitor at a certain location in the ring, and the orbit oscillations are clearly visible. By counting (or fitting) the number of oscillations in transverse planes, we obtain values of

$$Q_x = 64.31 , \quad Q_y = 59.32 . \quad (17.24)$$

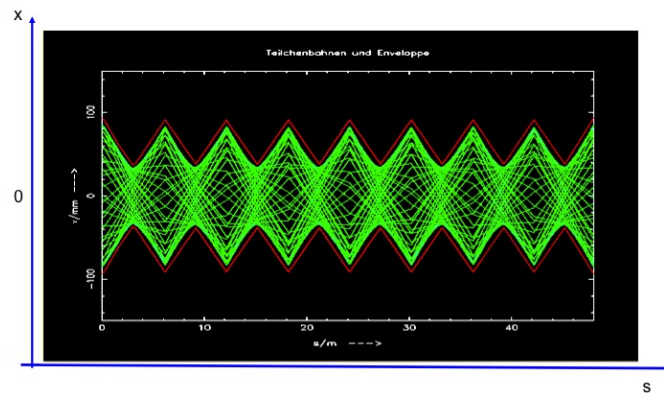
These values, which describe the eigenfrequencies of the particles, are called the horizontal and vertical *tune*, respectively. Knowing the revolution frequency, we can easily calculate the transverse oscillation frequencies, which for this type of machine usually lie in the range of  $\sim$ kHz.

As the tune characterizes the particle oscillations under the influence of external fields, it is one of the most important parameters of the storage ring. Therefore it is usually displayed at all times by the control system of such a machine. As an example, see the tune diagram of the HERA proton ring;



It was obtained via a Fourier analysis of the spectrum measured from the signal of the complete particle ensemble. The peaks indicate the two tunes in the horizontal and vertical planes of the machine, and in a sufficiently linear machine a fairly narrow spectrum is obtained.

Let us ask the question what the trajectory of the particle will look like for the second turn, or the third, or after an arbitrary number of turns. As we are dealing with a circular machine, the amplitude and angle,  $x$  and  $x'$ , at the end of the first turn will be the initial conditions for the second turn, and so on. After many turns the overlapping trajectories begin to form a pattern, which indeed looks like a beam:



There are a larger and smaller beam size but still remaining well-defined in its amplitude by the external focusing forces.

A mathematical function, which we call  $\beta$  or amplitude function, can be defined that describes the envelope of the single-particle trajectories. With this new variable, we can rewrite the equation for the amplitude of a particle's

transverse oscillations as

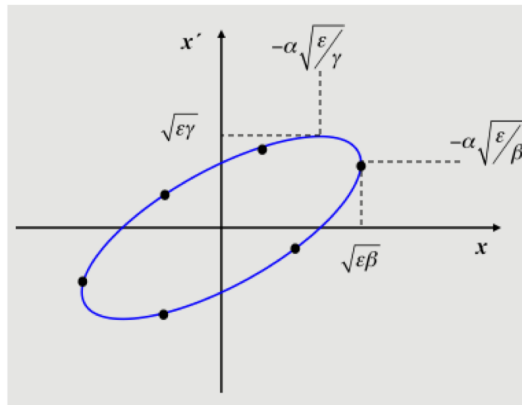
$$x(s) = \sqrt{\epsilon \cdot \beta(s)} \cos(\psi + \phi) , \quad (17.25)$$

where  $\psi(s)$  is the phase of the oscillation,  $\phi$  is its initial condition, and  $\epsilon$  is a characteristic parameter of the single particle or, if we are considering a complete beam, of the ensemble of particles. Indeed,  $\epsilon$  describes the space occupied by the particle in the transverse (here simplified two-dimensional)  $(x, x')$  phase space. More specifically, the area in  $(x, x')$  space that is covered by the particle is given by

$$A = \pi \cdot \epsilon , \quad (17.26)$$

and, as long as we consider conservative forces acting on the particle, this area is constant according to Liouville's theorem. Here we take these facts as given, but we point out that, as a direct consequence, the so-called emittance  $\epsilon$  cannot be influenced by whatever external fields are applied; it is a property of the beam, and we have to take it as given and handle it with care.

To following the usual textbook treatment of accelerators, we can draw in phase space the ellipse of the particle's transverse motion:



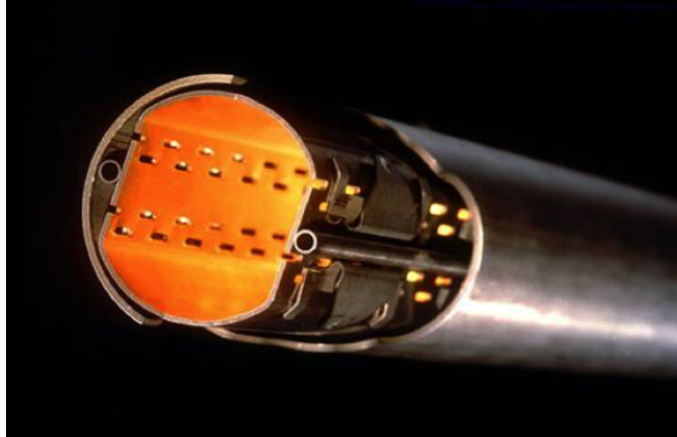
While the shape and orientation are determined by the optics function  $\beta$  and its derivative,  $\alpha = -\frac{1}{2}\beta'$ , the area covered is constant.

Let us talk a bit more about the beam as an ensemble of many (typically  $10^{11}$ ) particles. Referring to (17.25), at a given position in the ring the beam size is defined by the emittance  $\epsilon$  and the function  $\beta$ . Thus, at a certain moment in time the cosine term in (17.25) will be 1 and the trajectory

amplitude will reach its maximum value. Now, if we consider a particle at one standard deviation ( $\sigma$ ) of the transverse density distribution, then using the emittance of this reference particle we can calculate the size of the complete beam, in the sense that the complete area (within  $1\sigma$ ) of all particles in  $(x, x')$  phase space is surrounded (and thus defined) by our  $1\sigma$  candidate. Thus the value  $\sqrt{\epsilon \cdot \beta(s)}$  will define the  $1\sigma$  beam size in the transverse plane.

As an example, we use the values for the LHC proton beam: in the periodic pattern of the arc, the beta function is  $\beta = 180$  m and the emittance at flat-top energy is roughly  $\epsilon = 5 \times 10^{-10}$  rad m. The resulting typical beam size is therefore 0.3 mm.

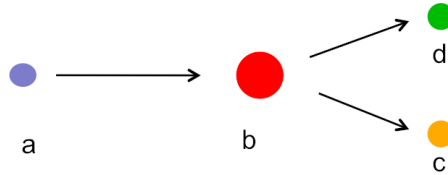
Clearly we would not design a vacuum aperture of the machine based on a  $1\sigma$  beam size. Typically, an aperture requirement is  $12\sigma$ , allowing for tolerances from magnet misalignments, optics errors and operational flexibility. Below the LHC vacuum chamber is shown (including the beam screen used to protect the cold bore from synchrotron radiation), which corresponds to a minimum beam size of  $18\sigma$ .



## 17.5 Limit V: Energy of Fixed-Target Colliders

The easiest way to perform physics experiments with particle accelerators is to bang the accelerated beam onto a fixed target and analyze the resulting events. While nowadays in high-energy physics we do not apply this

technique any more, it still plays an essential role in the regime of atomic and nuclear physics experiments. The advantage is that it is quite simple once the accelerator has been designed and built, and the particles produced are easily separated due to the kinematics of the reaction. The situation is illustrated below.



The particle "a" that is produced and accelerated in the machine is directed onto the particle "b", which is at rest in the laboratory frame. The particles produced from this collision are labelled "c" and "d" in this example.

While the set-up of such a scheme is quite simple, it is worth taking a closer look at the available energy in CM system. Considering the system of two particles colliding, we can write

$$(E_a^{\text{cm}} + E_b^{\text{cm}})^2 - (p_a^{\text{cm}} + p_b^{\text{cm}})^2 c^2 = (E_a^{\text{lab}} + E_b^{\text{lab}})^2 - (p_a^{\text{lab}} + p_b^{\text{lab}})^2 c^2 . \quad (17.27)$$

In the CM system we get, by definition,

$$p_a^{\text{cm}} + p_b^{\text{cm}} = 0 , \quad (17.28)$$

while in the Lab frame where particle "b" is at rest we have simply

$$p_b^{\text{lab}} = 0 . \quad (17.29)$$

The equation for the invariant mass therefore simplifies to

$$W^2 = (E_a^{\text{cm}} + E_b^{\text{cm}})^2 = (E_a^{\text{lab}} + m_b \cdot c^2)^2 - (p_a^{\text{lab}} \cdot c)^2 . \quad (17.30)$$

In other words, the energy that is available in the CM system depends on the square root of the energy of particle "a", which is the energy provided by the particle accelerator:

$$W \approx \sqrt{2E_a^{\text{lab}} \cdot m_b \cdot c^2} , \quad (17.31)$$

a quite unsatisfactory situation!

To meet the demand for higher and higher energies in particle collisions, the design of modern high-energy accelerators has naturally concentrated on the development of particle colliders, where two counter-rotating beams are brought into collision at one or several interaction points. See schematic diagram of the collision of two particles with equal energy:



If we calculate the available energy in the CM system for the case of two colliding beams of identical particles, we get

$$(p_a^{\text{cm}} + p_b^{\text{cm}})^2 = 0 \quad (17.32)$$

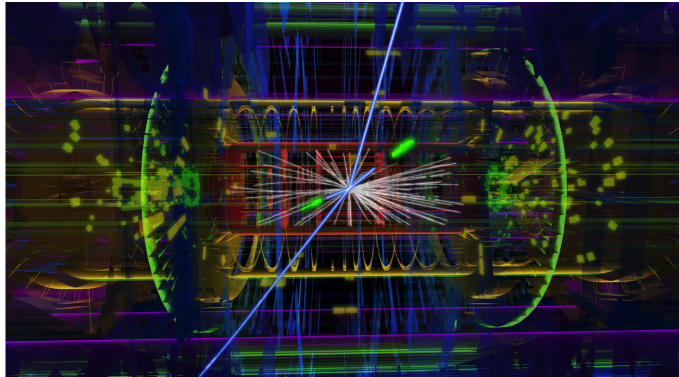
and, by symmetry, also

$$(p_a^{\text{lab}} + p_b^{\text{lab}})^2 = 0 . \quad (17.33)$$

So the full energy delivered to the particles in the accelerator is available during the collision process:

$$W = E_a^{\text{lab}} + E_b^{\text{lab}} = 2 E_a^{\text{lab}} . \quad (17.34)$$

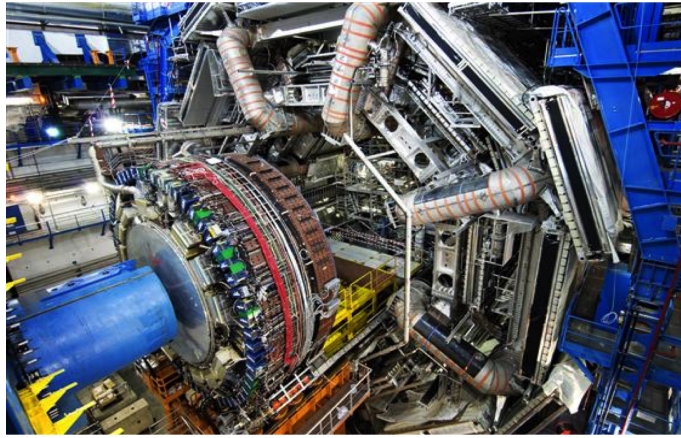
A ‘typical’ example of a high-energy physics event in such a collider ring (a Higgs particle measured in the ATLAS detector) is shown below.



## 17.6 Limit VI: The Unavoidable Particle Detectors

While it is quite clear that a particle collider ring is a magnificent machine in the quest for higher energies, there is a small problem involved, namely the

instalations of particle detectors. In the arc of the storage ring we can usually find a nice pattern of magnets providing us with a well-defined beam size, expressed as the beta function. However, special care has to be taken when our colleagues from high-energy physics wish to install a particle detector. Especially when working at the energy frontier, just like for the accelerators, these devices tend to expand considerably in size with the energies required. Below the largest particle detector installed in a storage ring is shown as an impressive example: the ATLAS detector at the LHC, which is 46 m in length and has an overall weight of 7000 t.



The storage ring has to be designed to provide the space needed for the detector hardware and at the same time create the smallest achievable beam spots at the collision point, which is usually right in the center of the detector. Unfortunately these requirements are a bit contradictory. The equation for the luminosity of a particle collider depends on the stored beam currents and the transverse spot size of the colliding beams at the interaction point:

$$L = \frac{1}{4\pi e^2 f_0 b} \cdot \frac{I_1 I_2}{\sigma_x^* \sigma_y^*} . \quad (17.35)$$

At the same time, however, the beta function in a symmetric drift grows quadratically as a function of the distance between the beam waist and the first focusing element, i.e.

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*} . \quad (17.36)$$

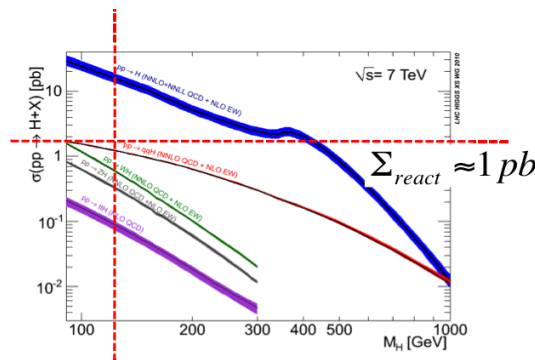
The smaller the beam at the interaction point, the faster it will grow until we can apply – outside of the detector region – the first quadrupole lenses.



As a consequence, this trend sets critical limits on the achievable quadrupole aperture or, for a given aperture, the achievable quadrupole gradient. The focusing lenses right before and after the interaction point, being placed as closed as possible to the detector, are generally the most critical and most expensive magnets in the machine, and their aperture requirement ultimately determines the luminosity that can be delivered by the storage ring.

### 17.7 Limit VII: The Rareness of Searched Events

The rate of events produced in a particle collision process depends not only on the performance of the colliding beams but first and foremost on the probability of creating such an event, the so-called cross-section of the process. In the case of the Higgs particle, which is without doubt the highlight of LHC Run 1, the overall cross-section is displayed below.



Without going into details, we can state that the cross-section for Higgs production is on the order of  $\Sigma_{\text{react}} \simeq 1 \text{ pb}$ . During the three years of LHC Run 1, i.e. the period 2011–2013, an overall luminosity of

$$\int L dt = 25 \text{ fb}^{-1} \tag{17.37}$$

was accumulated. Combining these two numbers using the fact that the event rate of a reaction is:

$$R = L \cdot \Sigma_{\text{react}} , \tag{17.38}$$

we get a total number of ‘some thousand’ Higgs particles produced – for a Nobel prize-winning event just at the edge of reliable statistics. Therefore,

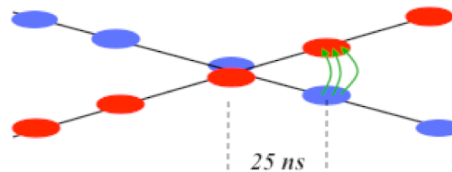
the particle colliders have to be optimized not only for the highest achievable energies but also for maximum stored beam currents and small spot sizes at the interaction points so as to optimize the luminosity of the machine.

## 17.8 Limit VIII: Luminosity of a Collider Ring

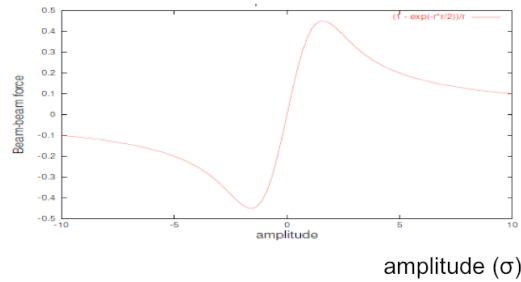
The design goal of a collider is to prepare, accelerate and store two counter-rotating particle beams in order to profit best from the energy of the two beams during the collision process. Still, there is a price to pay: unlike in fixed-target experiments, where the ‘particle’ density of the target material is extremely high, in the case of two colliding beams the event rate is basically determined by the transverse particle density that can be achieved at the interaction points. Assuming Gaussian density distributions in transverse planes, the performance of such a collider is described by the luminosity

$$L = \frac{1}{4\pi e^2 f_0 b} \cdot \frac{I_1 I_2}{\sigma_x^* \sigma_y^*} . \quad (17.39)$$

While the revolution frequency  $f_0$  and the bunch number  $b$  are ultimately determined by the size of the machine, the stored beam currents  $I_1$  and  $I_2$  and the beam sizes  $\sigma_x^*$  and  $\sigma_y^*$  at the interaction points have their own limitations. The most serious limitation comes from the beam–beam interaction itself. During the collision process, individual particles of the counter-rotating bunches feel the space charge of the opposing bunch. In the case of a  $pp$  collider, this strong field acts like a defocusing lens, and has a strong impact on the tune of bunches. The situation is shown schematically below.

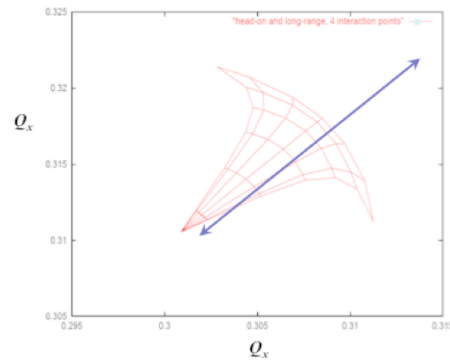


Two bunch trains collide at the interaction points, and during the collision process a direct beam–beam effect is observed. In addition to that, before and after the actual collision, long-range forces exist between bunches that have a nonlinear component; see the figure:



As a consequence, the tune of the beams is not only shifted with respect to the natural tune of the machine but also spread out, as different particles inside bunches see different contributions from the beam-beam interaction.

Therefore, in the tune diagram, we no longer obtain a single spot representing the ensemble of particles, but rather a large array that depends in shape, size and orientation on the particle densities, the distance of the bunches at the long-range encounters, and the single-bunch intensities. The effect has been calculated for the LHC and is displayed in the diagram:



In a number of cases a useful approximation can be applied, as for distances of about  $1 - 2 \sigma$ . Accordingly, a tune shift can be calculated to characterize the strength of the beam-beam effect in a collider. Given the parameters described above, and introducing the classical particle radius  $r_p$ , the amplitude function  $\beta^*$  at the interaction points and the Lorentz factor  $\gamma$ , we can express the tune shift due to the linearized beam-beam effect as

$$\Delta Q_y = \frac{\beta_y^* \cdot r_p \cdot N_p}{2\pi \gamma (\sigma_x + \sigma_y) \sigma_y}. \quad (17.40)$$

In the case of the LHC, the design value of the beam-beam tune shift is  $\Delta Q = 0.0033$ , and in daily operation the machine is optimized to run close to this value, which places the ultimate limit on achievable bunch intensities in the collider.

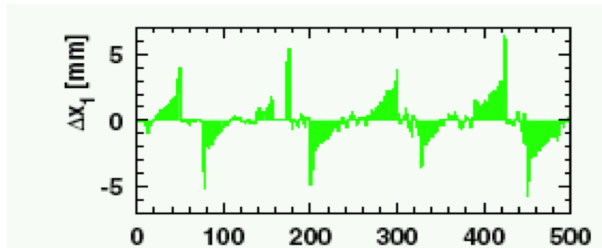
## 17.9 Limit IX: Power Loss in Synchrotron Radiation

In proton or heavy-ion storage rings, the design can more or less follow the rules discussed above. But the situation changes drastically as the particles become more and more relativistic. Bent on a circular path, electrons in particular will radiate an intense light, the so-called synchrotron radiation, which will have a strong influence on the beam parameters as well as on the design of the machine.

Summarizing the situation briefly here, the power loss due to synchrotron radiation depends on the bending radius and the energy of the particle beam:

$$P_s = \frac{2}{3} \alpha \hbar c^2 \frac{\gamma^4}{\rho^2}, \quad (17.41)$$

where  $\alpha$  represents the fine structure constant and  $\rho$  the bending radius in the dipole magnets of the ring. As a consequence, the particles will lose energy turn by turn. To compensate for these losses, RF power has to be supplied to the beam at any moment. An example that illustrates the problem nicely is shown in the figure:

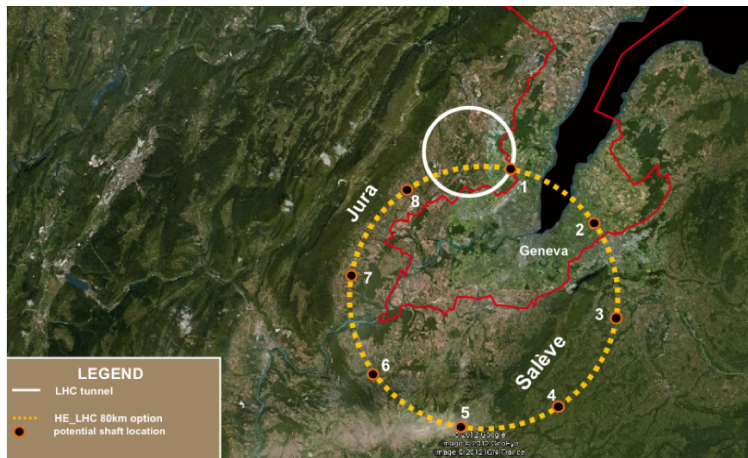


It plots the horizontal orbit of the former Large Electron–Positron Collider (LEP) storage ring. The electrons, travelling from right to left in the plot, lose a considerable amount of energy in each arc and hence deviate from the

ideal orbit towards the inner side of the ring. The effect on the orbit is large: up to 5 mm orbit deviation was observed. In order to compensate for these losses, four RF stations were installed in the straight sections of the ring to supply the necessary power.

The strong dependence of the synchrotron radiation losses on the relativistic  $\gamma$  factor sets severe limits on the beam energy that can be carried in a storage ring of a given size. The push for ever higher energies means either that storage rings even larger than LEP need to be designed or, to avoid synchrotron radiation, linear accelerating structures should be developed.

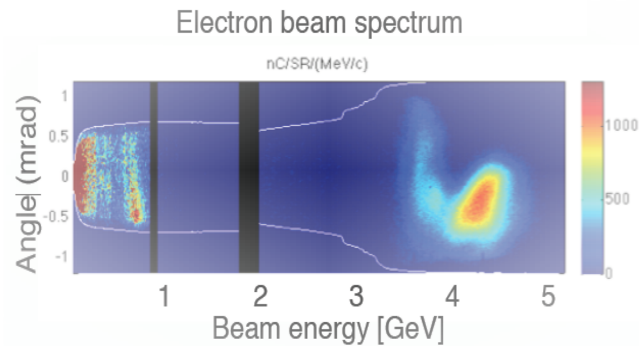
Currently, the next generation of particle colliders is being studied. The ring design of the future circular collider (FCC) in the Geneva region foresees a 100 km ring to carry electrons (and positrons) of up to 175 GeV energy. The size of this storage ring is far beyond the dimensions of anything that has been designed up to now. A sketch of the machine layout is given below, where the yellow dashed circle delineates the 100 km ring and the white circle represents the little LHC machine.



For the maximum projected electron energy of  $E = 175$  GeV, synchrotron radiation would cause an energy loss of 8.6 GeV, or an overall power of 47 MW of the radiated light at full beam intensity.

To summarize, for future lepton ring colliders synchrotron radiation losses set a severe limit on the achievable beam energy; and very soon the size of the machines will become uneconomical. For a given limit in synchrotron radia-

tion power, the dimensions of the machine would have to grow quadratically with the beam energy. Linear colliders are therefore proposed as the preferred way to go. In this case, the maximum achievable acceleration gradient is the key issue. New acceleration methods, namely plasma-based set-ups in which gradients have been observed that are much higher than those seen with conventional techniques, are a most promising concept for the design of future colliders. An impressive example is shown below:

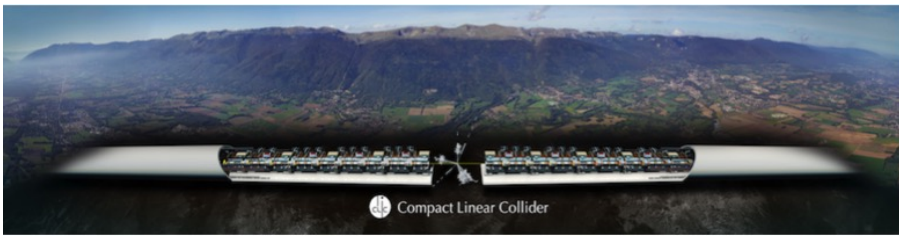


within a plasma cell of only a few centimeters in length, electrons are accelerated to several GeV. The gradients achievable are orders of magnitude higher than in any conventional machine. Still, there are problems to overcome, such as issues with overall efficiency, beam quality (mainly the energy spread of the beam), and the achievable repetition rate. Nevertheless, we are convinced that this is a promising field worthy of much further study.

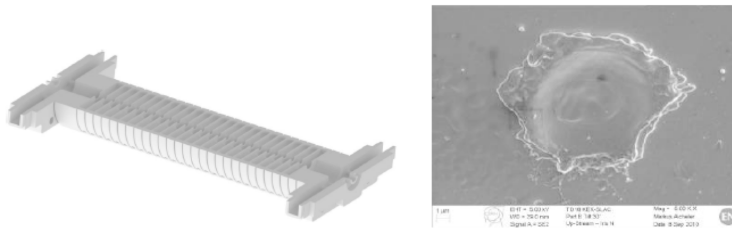
## 17.10 Limit X: Gradients in Linear Structures

Lepton colliders suffer from the severe limitation caused by synchrotron radiation losses, and at a certain point the construction of large facilities would not seem reasonable. To avoid the problem of synchrotron radiation, linear structures that were discussed earlier and used in the infancy of particle accelerators have become in vogue again. Still, the advantage of circular colliders cannot be completely ignored: even with a modest acceleration gradient in the RF structures, the particles will get turn by turn a certain boost in energy and will at some point reach the desired flat-top energy in the ring. In a linear accelerator, this kind of repetitive acceleration is by

design not possible; within a single pass through the machine, the particles will have to be accelerated to full energy. In order to keep the structure compact, the highest acceleration gradients will therefore be needed. One of the most prominent designs proposed for a possible future collider is the CLIC design, the linear collider along the Jura Mountains in the Geneva region. Within one passage through the 25 km long accelerator, the electron beam will get up to 3 TeV, and the same is true for the opposing positron beam. An artist's rendering of this machine is shown below.



A picture of the CLIC test facility CTF3 is shown below.



On the right-hand side is an electron microscope photo of the surface after a voltage breakdown. A little crater can be seen, indicating possible damage to the surface and, as a consequence, deterioration of the achievable gradient which has to be avoided under all circumstances. Although considerably higher than the typical values in circular machines, the gradient  $E_{\text{acc}} = 100 \text{ MV m}^{-1}$  in a linear machine still leads to a design of overall length approximately 50 km for a maximum achievable energy of  $E_{\text{max}} = 3 \text{ TeV}$ .

The main parameters of the CLIC design are listed in Tab. 17.1. The accelerating gradient, i.e. the energy gain per meter, is especially to be emphasized; it has been pushed to the maximum value that is technically feasible, and the limit is ultimately due to the breakdown of the electric field in the accelerating structure.

Table 17.1: Main parameters of the CLIC study

	500 GeV	3 TeV
Site length	13 km	48 km
Loaded acceleration gradient ( $\text{MV m}^{-1}$ )	12	
Beam power per beam (MW)	4.9	14
Bunch charge ( $10^9$ e+/e)	6.8	3.7
Horiz./vert. emittance ( $10^{-6}/10^{-9}$ m)	2.4/25	0.66/20
Beta function (mm)	10/0.07	
Beam size at IP: horizontal/vertical (nm)	45/1	
Luminosity ( $\text{cm}^{-2} \text{s}^{-1}$ )	$2.3 \times 10^{34}$	$5.9 \times 10^{34}$

**Exercise 17.1:** The Princeton synchrotron has been used to accelerate highly charged nitrogen ions. If this synchrotron can produce protons of total energy  $3 \text{ GeV}$ , what is the maximum kinetic energy of  $^{14}\text{N}$  ions?

**Exercise 17.2:** A certain  $e^-e^+$  pair produced cloud chamber tracks of radius of curvature  $3 \text{ cm}$  in a plane perpendicular to the applied magnetic field  $0.11 \text{ Tesla}$ . What was the energy of the  $\gamma$ -ray which produced the pair?

**Exercise 17.3:** A proton and a electron with the total energy  $1.4 \text{ GeV}$  each, transverse two scintillation counters  $10 \text{ m}$  apart. What are the times of flight of these particles?

**Exercise 17.4:** Briefly describe the cyclotron and the synchrotron, contrasting them. Tell why one does not use: cyclotrons to accelerate protons to  $2 \text{ GeV}$  and synchrotrons to accelerate electrons to  $30 \text{ GeV}$ ?

**Exercise 17.5:** A relativistic proton loses  $1.8 \text{ MeV}$  when penetrating a  $1 \text{ cm}$  thick scintillator. What is the most likely mechanism: (a) Ionization, excitation; (b) Compton effect; (c) Pair production.

**Exercise 17.6:** The efficiency of a proportional counter for charged particles is ultimately limited by: (a) Signal-to-noise ratio; (b) Total ionization; (c) Primary ionization.

**Exercise 17.7:** The light emission in organic scintillators is caused by transitions between: (a) Levels of delocalized electrons; (b) Vibrational levels; (c) Rotational levels.



## Chapter 18

# History of Validation of SM

In 1967, Weinberg published a paper where he stated: "Leptons interact only with photons, and with the intermediate bosons that presumably mediate weak interactions. What could be more natural than to unite these spin-one bosons into a multiplet of gauge fields"? Most of the ingredients of what would become the SM were in place in the early 1970's, and in 1971–72 t'Hooft and Veltman showed that the theory was renormalizable.

### 18.1 Discovery of Neutral Currents

A stunning feature of the SM was that it predicted a new interaction: the weak *Neutral Current* (NC). This was the first time a fundamental interaction was predicted before it was observed. It was clearly a triumph in 1981 to see  $W$ 's and  $Z$ 's directly, but it was the observation of the NC that convinced most physicist that the SM was right. The NC couplings are entirely decided by the (anti)fermion charge and  $\sin\theta_W$ . Neutrino scattering with nuclei offer possibility of studying NC interactions of quarks. These typically have higher event rates compared to the pure leptonic scattering processes due to the possibility of using nuclear targets. However, analysis of these processes requires an understanding and knowledge of the proton structure. Hence the cleanest probe of the NC couplings can come from analyzing pure leptonic reactions.

The easiest way to look for evidence of NC is to make  $Z$ 's directly ( $e^+e^- \rightarrow$

$Z^0$ ) or ( $q\bar{q} \rightarrow Z^0$ ). However, there was no machine capable of doing that in the early 1970's. The mass of the  $Z^0$  is approximately 92 GeV. None of the machines available in the 1970's could produce 92 GeV in the CM system! Some of the experimental facilities operating in the early 1970's were:

- **SLAC:** A linear accelerator that could produce a 22 GeV  $e^-$  beam. They were also just turning on an  $e^+e^-$  storage ring SPEAR with a maximum energy of  $2.6 \times 2.6$  GeV (5.2 GeV in the CM).
- **FNAL:** Just turning on with 200 GeV  $p$  beam (increased to 400 GeV by end of the decade). 200 GeV  $p$  on a fixed target gives approximately 20 GeV in the CM ( $E_{\text{cm}} = \sqrt{2E_{\text{lab}}m}$ ) so again they were not able to produce  $Z$  bosons directly.
- **CERN:** A proton synchrotron produced a 28 GeV  $p$  beam which could be used to make a  $28 \times 28$  GeV  $pp$  collider (ISR). In the late 1970's, CERN upgraded the ISR to a  $270 \times 270$  GeV  $p\bar{p}$  storage ring which is where the  $Z$  was directly produced and detected for the first time.
- **BNL** A 33 GeV  $p$  beam on fixed target.

To study the properties of the weak NCs it was necessary first to establish its existence. To that end, it was necessary to predict the characteristics of the events that would result from interactions of  $\nu_\mu$ ,  $\bar{\nu}_\mu$  and  $\bar{\nu}_e$  beams with electrons, as that would be the cleanest probe.

Let us list different types of elastic scattering processes involving just leptons that can take place through weak *Charged Current* (CC) and NC interactions using the  $\nu$  beams and the electron targets. These are:

1.  $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ , which can take place only through the CC interaction;
2.  $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$  and  $\bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^-$ , which can take place only through the NC interaction;
3.  $\nu_e + e^- \rightarrow \nu_e + e^-$ ;  $\bar{\nu}_e + e^- \rightarrow \bar{\nu}_e + e^-$ , which can take place both through the NC and CC interactions.

Calculation of the scattering amplitudes of various NC and CC processes listed above proceeds using the usual rules of QFT. For the low energies of

$\nu$ -beams that were available then, the  $M_W, M_Z \rightarrow \infty$  approximation could be used.

Table 18.1 shows the differential cross-section in terms of the variable  $y = E_e/E_\nu$  and the integrated cross-section.

Process	$d\sigma/dy$	$\sigma$
$\nu_\mu + e^- \rightarrow \mu^- + \nu_e$	$A s (g_L^\nu)^2 (g_L^e)^2$	$A s (g_L^\nu)^2 (g_L^e)^2$
$\nu_\mu + e^- \rightarrow \nu_\mu + e^-$	$A s (g_L^\nu)^2 [(g_L^e)^2 + (1-y)^2 (g_R^e)^2]$	$A s (g_L^\nu)^2 [(g_L^e)^2 + \frac{1}{3} (g_R^e)^2]$
$\bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^-$	$A s (g_L^\nu)^2 [(g_R^e)^2 + (1-y)^2 (g_L^e)^2]$	$A s (g_L^\nu)^2 [\frac{1}{3} (g_L^e)^2 + (g_R^e)^2]$
$\nu_e + e^- \rightarrow \nu_e + e^-$	$A s (g_L^\nu)^2 [(g_L^e + 1)^2 + (1-y)^2 (g_R^e)^2]$	$A s (g_L^\nu)^2 [\frac{1}{3} (g_R^e)^2 + (g_L^e + 1)^2]$
$\bar{\nu}_e + e^- \rightarrow \bar{\nu}_e + e^-$	$A s (g_L^\nu)^2 [(g_R^e)^2 + (1-y)^2 (g_L^e + 1)^2]$	$A s (g_L^\nu)^2 [\frac{1}{3} (g_L^e + 1)^2 + (g_R^e)^2]$

Table 18.1: The differential and total cross-sections for a few  $\nu, \bar{\nu}$  induced CC and NC processes, with  $A = 4G_\mu^2/\pi$ .

A few comments are in order. The above expressions use  $\rho = 1$  as well as the fact that values of  $g_L, g_R$  for the  $\mu$  and the  $e$  are the same. All the neutrino induced cross-sections are indeed proportional to the square of the CM energy  $s$ . The variable  $y$  is related to the scattering angle  $\theta$  in the CM frame. One can see after some manipulations that the angular distribution of the scattered charged lepton is different for the case of  $\nu$  and  $\bar{\nu}$ . In the first row we have written the cross-section for the CC process:  $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$ ,

so that one can indeed see that the size of the expected cross-sections for the NC processes are of the same order of magnitude as the CC process and depend on  $\sin \theta_W$ .

Note the last two rows of Tab. 18.1. As one changes from the  $\nu_\mu, \bar{\nu}_\mu$  beams to  $\nu_e, \bar{\nu}_e$  beams the factors of  $(g_L^e)^2$  in the total cross-section expressions gets changed to  $(g_L^e + 1)^2$ . Further, note also the different weights of the  $(g_L^e)^2$  and  $(g_R^e)^2$  contributions as one changes from  $\nu$  to  $\bar{\nu}$  beams.

Both these observations tell us that the contours of constant cross-section for these four processes are ellipses in the  $g_A - g_V$  plane with different centers and with major axes of differing orientations. Thus a measurement of these cross-sections can then help us determine  $g_V^e, g_A^e$ , albeit up to sign ambiguities.

Note also from the table that as one changes from  $\nu$  to  $\bar{\nu}$ , the terms in the angular distribution proportional to  $(g_L^e)^2$  and  $(g_R^e)^2$  get interchanged. This behavior can be understood very easily in terms of the chirality conservation of the gauge interaction and the angular momentum conservation. As a result, one can write the weak NC cross-sections for all the different pairs of fermions rather easily by inspection. In particular, the same table can be used to calculate the cross-section for the weak NC induced processes with nucleon (nuclear) targets as well. The hadronic weak NC events arise from the scattering of the  $u, d, s$  quarks in the nucleon (nucleus). In the parton model the net rate is then given by the incoherent sum over all the quarks contained in the nucleon (nucleus). Using the information on the momentum distributions of quarks/antiquarks in the nucleon (nucleus), it is also possible to estimate the expected cross-section. Again these too depend only on  $\sin^2 \theta_W$  as far as the EW model parameters are concerned.

At the time of the discovery of weak NCs in hadronic and leptonic production, theoretical estimates were available for the upper limit on the ratio of NC to CC elastic scattering. This was obtained by using experimental knowledge of the form factor of the proton and neutron. The same was also available for the inelastic process of the inclusive production of hadrons using the language of structure functions of the target nucleus. Two points are worth noting here. While the use of nuclear targets increased the expected rates for NC induced hadron production, establishing that the events are indeed due to weak NC was difficult because of the large neutron induced background. The pure leptonic processes on the other hand, were predicted to be very rare and hence difficult to observe, but could unambiguously prove existence of weak NC as soon as even one event was observed.

NCs were discovered in 1973 in the study of elastic scattering of  $\nu_\mu$  and  $\bar{\nu}_\mu$  off nuclear targets. The experiment discovered evidence for the NC induced hadronic processes

$$\nu_\mu + N \rightarrow \nu_\mu + \text{hadrons} ; \quad \bar{\nu}_\mu + N \rightarrow \bar{\nu}_\mu + \text{hadrons} . \quad (18.1)$$

as well as pure leptonic processes,

$$\bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^- , \quad (18.2)$$

using the giant bubble chamber Gargamelle. Below you can see the photo of observation of the first leptonic interaction induced by weak NC. In fact the discovery came in an experiment which had been designed to study the CC interactions:

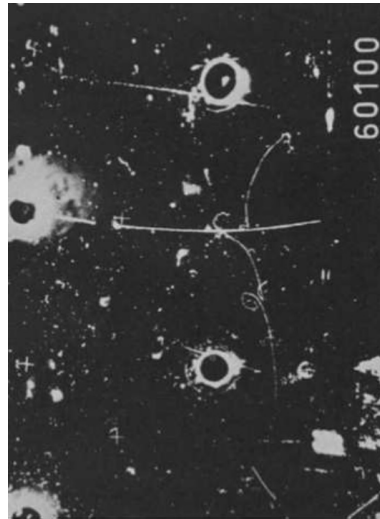
$$\nu_\mu + N \rightarrow \mu^- + \text{hadrons} ; \quad \bar{\nu}_\mu + N \rightarrow \mu^+ + \text{hadrons} . \quad (18.3)$$

Thus the experiment could easily extract the ratio of the CC to NC events, after the observation of NC in hadronic events. The experiment had seen  $\sim 100$  events of different categories (NC and CC) containing hadrons, with

$$\left. \frac{NC}{CC} \right|_\nu = 0.21 \pm 0.03 ; \quad \left. \frac{NC}{CC} \right|_{\bar{\nu}} = 0.45 \pm 0.09 . \quad (18.4)$$

As already mentioned, the same experiment also found evidence for the pure leptonic process, where the  $\nu_\mu$  was scattered off the atomic electron.

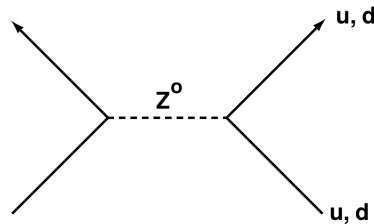
The figure below shows the image of the first unambiguous, weak NC event ever observed.



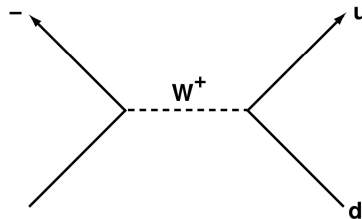
The incoming antineutrino, interacts with an atomic electron and knocks it forward. The electron is identified from the characteristic shower created by the electron-positron pairs. This was considered to be clear evidence for the weak NC. The theoretical predictions summarized in Tab. 18.1 were used to justify the interpretation. With just one  $\bar{\nu}$  event, the experiment could only quote a range  $0.1 < \sin^2 \theta_W < 0.6$  at 90% c.l. The number of hadronic NC events on the other hand, was big enough to extract a value of  $\sin^2 \theta_W$  to be in the range of 0.3 - 0.4. This was the first qualitative validation of the prediction of NCs.

As we mentioned above no machine could produce  $Z$  bosons directly in the early 1970's and the first observation of NC had to be indirect. The most sensitive technique was neutrino scattering, where a  $\nu_\mu$  scattered off the quarks in a nuclear target. The hadrons in the final state could be detected (the incoming and outgoing  $\nu$  were invisible to the detector), and the absence of a muon meant it was a NC event. The rate for the NC process could be compared to the rate for the corresponding CC process where in addition to the hadrons from the breakup of the nucleon, an accompanying muon could be seen. Below you can see the diagrams for NC (a), and CC (b), neutrino scattering.

(a) Neutral Current:



(b) Charged Current



It is straightforward to work out the cross sections for  $\nu$  and  $\bar{\nu}$  scattering off nucleons. The ratios of cross sections are what can be experimentally

measured most precisely and the experimentally accessible quantities are:

$$\begin{aligned} R_\nu &\equiv \frac{\sigma(\nu N \rightarrow \nu X)}{\sigma(\nu N \rightarrow \mu^- X)} = \frac{1}{2} - \sin^2 \theta_W + \frac{20}{27} \sin^4 \theta_W, \\ R_{\bar{\nu}} &\equiv \frac{\sigma(\bar{\nu} N \rightarrow \bar{\nu} X)}{\sigma(\bar{\nu} N \rightarrow \mu^+ X)} = \frac{1}{2} - \sin^2 \theta_W + \frac{20}{9} \sin^4 \theta_W. \end{aligned} \quad (18.5)$$

Measuring absolute cross sections involves a detailed knowledge of the flux of the incident  $\nu$  beam and that is hard to know; however, in the ratio, both the flux and the poorly known energy spectrum of the  $\nu$  beam cancel.

By just seeing  $\nu N \rightarrow \nu X$  reactions, one observes NC for the first time which is a great achievement; however, one can also use the cross section ratio to extract  $\sin^2 \theta_W$  (or whatever your favorite third  $SU(2) \times U(1)$  parameter is; the convention was to use  $\sin^2 \theta_W$  until LEP came on line and now  $M_Z$  is standard). Once  $\sin^2 \theta_W$  is known, all the SM couplings are determined and by measuring  $R_\nu$  and  $R_{\bar{\nu}}$  one gets a wonderful consistency check. If both give the same value of  $\sin^2 \theta_W$ , it is an indication one has chosen the right gauge structure (for example,  $SU(2)_L \times SU(2)_R \times U(1)$  would predict different relations between  $R_\nu$  and  $R_{\bar{\nu}}$ ) and one is starting to test the predictive power of the theory.

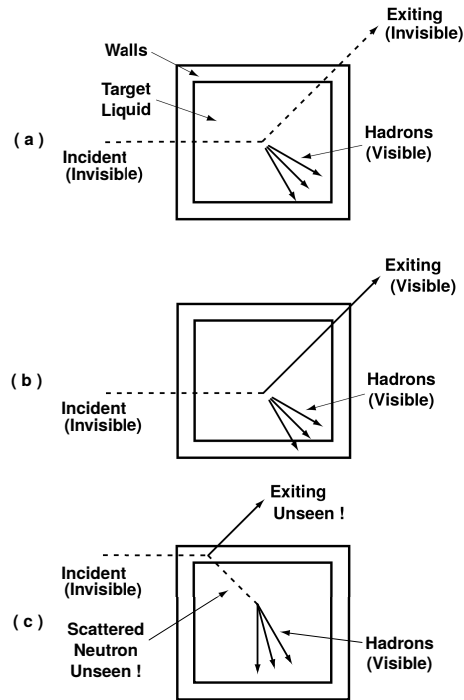
To make neutrinos, one starts with a proton beam on a target that produces lots of secondary particles; in particular, lots of kaons and pions will be produced. The kaons and pions are selected for sign and then allowed to decay, ( $\pi \rightarrow \mu\nu_\mu$ ,  $K \rightarrow \mu\nu_\mu$ ,  $K \rightarrow \pi\mu\nu_\mu$ ) and neutrinos are produced. Muon neutrinos are strongly favored by helicity, and neutrinos or antineutrinos are selected by the charge of the meson. The experiments are hard. The major obstacle is just rate. The  $\nu N$  scattering cross section is proportional to  $G_F^2 M_p E_\nu$  and  $G_F$  is a small number so the cross section is small.

$$\sigma_{CC}^{\nu N} \sim 6 \times 10^{-6} \frac{\text{nb}}{\text{nucleon}} \frac{E_\nu}{\text{GeV}}. \quad (18.6)$$

Working at the highest possible  $\nu$  beam energy is clearly an advantage. FNAL with its 200 GeV  $p$  beam had a great advantage for making high energy neutrinos over CERN with 30 GeV  $p$ , but CERN got there first.

The detector that made the discovery was the 12' Gargamelle bubble chamber. The central part of the detector was a big tank of supersaturated Freon. When a charged particle passed through the supersaturated gas, it left an ionization trail. The gas was expanded suddenly after the beam pulse and

bubbles formed along the ionization trail. The bubble tracks were then photographed, scanned, and measured by hand for evidence of interesting physics processes. An important feature of the detector was the ability to identify NC events by having good solid angle coverage for muons so a muon could not escape undetected.



The figure above illustrates what CC, NC and background events would look like in the detector. (a) – NC event. The incoming and escaping neutrinos are invisible and the signal is the observation of a hadronic cluster. (b) – CC event. The exiting muon is observed. (c) – Background event. An incoming neutrino interacts in the material of the chamber wall, producing a neutral hadron which cannot be detected and a muon that escapes. The neutral hadron can then interact in the chamber and look like a NC event. The most worrisome background was a  $\nu$  interacting in the material of the chamber wall, producing a neutral hadron and an escaping muon. The neutral hadron could then interact in the chamber and look like NC event. The experiment took some 300,000 pictures, 83,000 with the  $\nu$  beam and 207,000 with the  $\bar{\nu}$  beam. They collected twice as many events for the  $\bar{\nu}$  beam since the scattering cross section was approximately one third the  $\nu$  cross section due



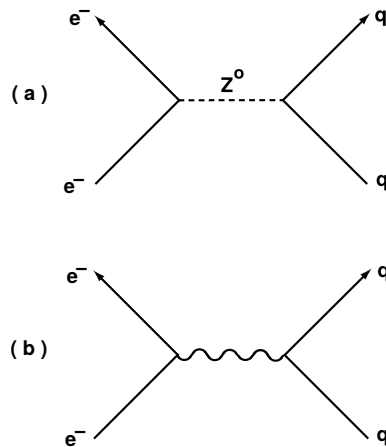
to helicity effects.

The experimenters spent a great deal of effort studying possible backgrounds to the NC sample from neutral hadrons. One particularly convincing check was that background from neutral hadrons was expected to show attenuation along the length of the chamber and they were able to show that their NC candidates had a uniform distribution along the chamber length. From the ratio of the number of NC to CC events, they were able to conclude " $\sin^2 \theta_W$  is in the range 0.3–0.4". They conservatively claimed "if the events are due to NC, then  $R_\nu$  and  $R_{\bar{\nu}}$  are compatible with the same value of  $\sin^2 \theta_W$ ". This experiment has been repeated many times since 1973.

## 18.2 The Discovery of Neutrinoless NC

NCs were discovered by neutrino scattering experiments. The coupling that was observed was consistent with the SM predictions; however, there are many other terms in the SM Lagrangian involving NC apart from neutrino-quark interactions. In particular, there are NC terms in the Lagrangian that do not involve neutrinos (e.g. electron-quark scattering through  $Z$  exchange). These terms are particularly interesting because electron-quark scattering can take place via  $Z$  or  $\gamma$  exchange and the two processes can interfere.

Below you can see the diagrams of the two processes: (a)  $Z$  exchange and (b)  $\gamma$  exchange, which can contribute to electron-quark NC scattering.



This interference allows one to explore the parity violating nature of the NC interaction.

There were two approaches that experimentalists used to probe the electron-quark coupling. The first approach was to scatter high-energy polarized electrons off of a nuclear target. This was first done at SLAC. The other was to use atoms. The  $e^-$  in an atom interacts with the nucleus both via the usual EM interaction and by  $Z$  exchange. The immediate consequence is that the atomic Hamiltonian does not conserve parity. Everything you learned about stationary states of atomic systems being eigenstates of the parity operator was incorrect (although a good approximation)!

In the late 1970's, experiments in atomic bismuth failed to detect parity violation, in contradiction with the SM expectation. The early atomic physics results were wrong. Later experiments in the early 1980's agreed with the SM predictions but by that time the SLAC experiment had already confirmed the SM predictions for electron-quark couplings in 1978–79.

The basic idea of the electron-quark scattering experiments is that the scattering cross section is the square of the sum of the weak and EM amplitudes,  $A_{WK}$  and  $A_{EM}$ , which can interfere:

$$\sigma \sim |A_{EM} + A_{WK}|^2 = A_{EM}^2 \left( 1 + \frac{2A_{EM}A_{WK}}{A_{EM}^2} + \frac{A_{WK}^2}{A_{EM}^2} \right). \quad (18.7)$$

At low  $Q^2$  ( $Q^2 < 10 \text{ GeV}^2$ ),  $A_{EM} \gg A_{WK}$  and the last term can be dropped. The interference term, however, can be detected.

If we treat the NC as current-current interaction with vector ( $V$ ) and axial vector ( $A$ ) parts (recall the CC is  $V - A$  but the NC is much more complicated) then

$$A_{WK} = J_e J_q = (V_e V_q + A_e A_q) + (V_e A_q + A_e V_q), \quad (18.8)$$

where the subscripts  $e$  and  $q$  refer to the electron and quark currents. The first term is a scalar,

$$(V_e V_q + A_e A_q) = A_{WK}^{\text{scal}}, \quad (18.9)$$

and is extraordinarily difficult to detect. The second term,

$$(V_e A_q + A_e V_q) = A_{WK}^{\text{ps-scal}}, \quad (18.10)$$

however, is a pseudoscalar and has a very nice signature because it changes sign under parity transformation. It is straightforward to show that if we

define the asymmetry,  $\delta$ , as the difference in the scattering cross section for left and right handed scattering divided by the sum, then:

$$\delta = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} = \frac{2(A_{WK}^{\text{ps-scal}} A_{EM})}{A_{EM}^2}, \quad (18.11)$$

where  $\sigma_R, \sigma_L$  are the cross sections for right and left handed coordinate systems and the handedness of the coordinate system is determined by, for example, the longitudinal polarization of the incoming  $e^-$  beam.

The asymmetry is small! At  $Q^2 \sim 10 \text{ GeV}^2$ , the ratio of the weak and electromagnetic amplitudes can be estimated:

$$\begin{aligned} A_{EM} &\sim \frac{4\pi\alpha_{EM}}{q^2}, \\ A_{WK} &\sim G_F, \\ \delta &\sim \frac{G_F q^2}{4\pi\alpha_{EM}} \sim 10^{-4}. \end{aligned} \quad (18.12)$$

In the SLAC experiment that discovered neutrinoless NC, high-energy polarized electrons were scattered off of an unpolarized deuterium target. The scattered electrons were detected at a fixed scattering angle in the lab which corresponds to a fixed energy of the scattered electron. It is straightforward (but tedious) to start from the SM Lagrangian and calculate the expression for the asymmetry in scattering left versus right handed electrons.

To measure an asymmetry of  $10^{-4}$  to 10% precision, one needs  $10^{10}$  events. Clearly one cannot count scattered electrons one by one. The experiment used a slightly different philosophy from the usual single particle counting techniques common in high energy physics. Instead of counting the scattered electrons individually, the detector integrated the signal and measured a current of scattered electrons on each beam pulse.

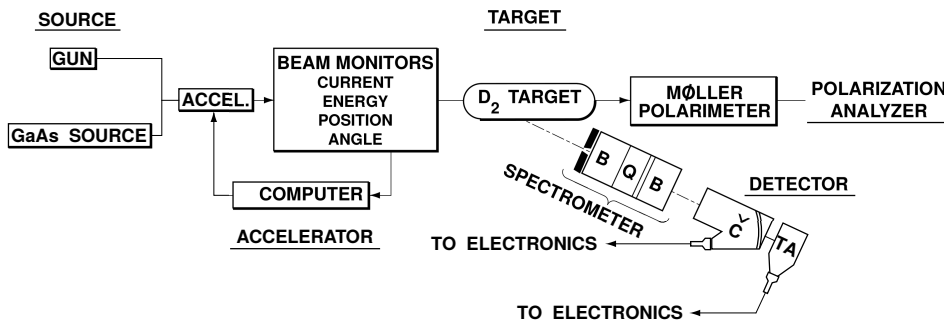
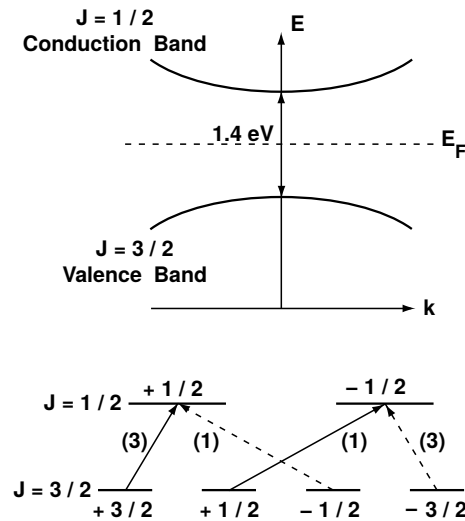


Figure above shows an overview of the SLAC polarized electron scattering experiment that discovered neutrinoless NC. At the gun end of the Linac, they started with a polarized  $e^-$  source. The polarized  $e^-$  source was really very cute. Usually in a linear accelerator, one uses a thermionic cathode which is heated up and electrons are boiled off, collected, and used to produce an unpolarized beam. To make polarized electrons, they replaced the thermionic cathode with a gallium arsenide crystal. The electrons were polarized by optically pumping electrons from the  $j = 3/2$  valance band to  $j = 1/2$  conduction band of the crystal with a circularly polarized laser beam (710 nm light). Starting from the valance band a circularly polarized photon has  $\Delta j_Z = +1$  or  $-1$ . The Clebsch-Gordan coefficients are favorable and one gets 3 times as many electrons in one  $m_j$  level as the other in the upper state conduction band, which polarizes the upper state. Below is illustrated schematically the energy levels of the conduction band and valance band of GaAs, a circularly polarized laser which induces transitions can polarize the upper state as shown.



To get the electrons out of the crystal conduction band, they coated the surface with cesium and oxygen which produced a negative work function. The electrons could escape and their polarization was preserved. The circular polarization of the laser controlled the polarization of the  $e^-$  beam and it could be changed on a pulse by pulse basis in a random way. This technique theoretically could produce an electron beam with 50% polarization. In practice, the average electron beam polarization was 37%.

The beam was then accelerated down the Linac with very little loss of polarization. At the end of the Linac, the beam was deflected into the beam switchyard onto the deuterium target.

The scattered  $e^-$  flux was measured with 2 independent detectors, both measuring the total charge passing through them. The polarization of the spent beam was determined with a Møller polarimeter, taking advantage of the asymmetry in the cross section for a longitudinally polarized electron scattering on polarized target electrons. The parity violating asymmetry that was the signature for NC was computed by counting electrons scattered into the detector when the electron beam was left handed versus right handed.

The challenge of this type of experiment is not just to measure an asymmetry of one part in  $10^4$ , but to convince that you are measuring the correct asymmetry! A great attention was paid to ensure that all possible instrumental asymmetries were at the  $10^{-5}$  level or smaller. The final result was

$$\sin^2 \theta_W = 0.222 \pm 0.018 . \quad (18.13)$$

It was a demonstration of the power of the SM that with a single value of the parameter  $\sin^2 \theta_W$ , it could account in detail for the strengths of very disparate processes: both the SLAC polarized electron scattering experiment and the neutrino scattering experiments. And that is why, with the SLAC result, the high energy community was convinced, even before the  $Z$  was found, that the SM was correct and that  $SU(2)_L \times U(1)$  was the correct gauge group to describe the world around us.

### 18.3 Observation of Charm

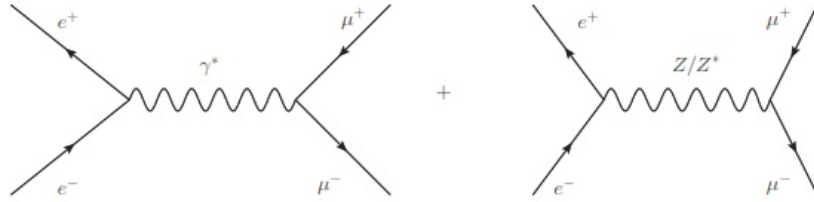
Soon after the observation of the weak NC, the charm quark was also discovered with mass very close to that predicted by the analysis of the  $\Delta S = 2$ ,  $K_0 - \bar{K}_0$  mixing caused by FCNC. As we understand now, in view of the very large mass of the top quark, it was somewhat 'fortuitous' that the charm quark contribution to the  $\Delta S = 2$  mass difference was the dominant one. Be as it may, this was an extremely important second validation of the correctness of the gauge theory of EW interactions based on the gauge group  $SU(2)_L \times U(1)_Y$ . Note that one of the validation came from tree level couplings and the other from loop induced effects.

### 18.4 Fixing of $\sin^2 \theta_W$ and Prediction of $M_W, M_Z$

The same leptonic couplings which contribute to the NC scattering processes involving  $\nu$ 's can also make their presence felt in processes like

$$e^+ + e^- \rightarrow \mu^+ + \mu^- . \quad (18.14)$$

This proceeds through a  $\gamma^*$  exchange in the  $s$ -channel and a  $Z/Z^*$  exchange shown below.



Whether the  $Z$  will be on shell or off shell of course depends on the CM energy.

The cross-section for this process can be easily computed. Electromagnetic interactions being the same for the left and right chiral fermions,  $\gamma^*$  exchange diagram gives a forward-backward symmetric contribution whereas both, the square of the amplitude of the  $Z$  exchange diagram itself and the interference term, will give contributions which are forward backward asymmetric. Hence the presence of the weak NC will manifest itself in the form of a forward-backward asymmetry in (say)  $\mu$  production. Both the size and sign of this asymmetry depends on the CM energy of the process  $\sqrt{s} = 2E_b$  where  $E_b$  is the beam energy, relative to the mass of the  $Z$  boson.

In fact if  $\theta$  is the angle made by the outgoing lepton with the incoming lepton, then one can show that

$$\frac{d\sigma(e^+e^- \rightarrow \mu^+\mu^-)}{d\cos\theta} = \frac{\pi\alpha_{\text{em}}^2}{2s} [A(1 + \cos^2\theta) + B\cos\theta] , \quad (18.15)$$

where

$$\begin{aligned} A &= 1 + 2\Re(\chi)g_V^2 + |\chi|^2(g_V^2 + g_A^2)^2 , \\ B &= 4\Re(\chi)g_A^2 + 8|\chi|^2g_V^2g_A^2 , \\ \chi &= \left( \frac{G_\mu M_Z^2}{2\sqrt{2}\pi\alpha} \right) \frac{s}{s - M_Z^2 + iM_Z\Gamma_Z} . \end{aligned} \quad (18.16)$$

Here  $g_V, g_A$  denote the vector and the axial vector NC couplings for  $e$  and  $\mu$ ,  $\Gamma_Z$  is the width of  $Z$ . In the chosen normalization, deviation of  $A$  from 1 and that of  $B$  from zero is then indication of the contribution of the weak NC to the process. Both  $A$  and  $B$  contain terms linear in  $\Re|\chi|$  and  $g_V^2$  or  $g_A^2$ . Hence, even for small values of  $|\chi|$ , both the total cross-section and the angular distribution can be used to probe the weak NC contribution.  $B$  is zero without the  $Z$  contribution. It is however nonzero for both, the interference terms containing  $\Re(\chi)$  and the square of the  $Z$ -exchange diagram alone, containing  $|\chi|^2$ . Hence the angular distribution contains an asymmetric term at all  $s$ . If we analyze these expressions we find that the results for this asymmetry are very different for  $\sqrt{s} \ll M_Z$  and  $\sqrt{s} = M_Z$ .

The forward-backward asymmetry,  $A_{FB}^\mu$  defined as the ratio of the difference of cross-sections with the  $\mu^-$  in the forward and the backward hemisphere and the total cross-section, is then expected to be nonzero due to the  $Z$  contribution and the interference term. It is clear that this is also the same as charge asymmetry between the muons in the forward hemisphere. Thus one has two asymmetries  $A_{FB}^\mu$  and  $A_C^\mu$ :

$$\begin{aligned} A_{FB}^\mu &= \frac{\sigma(\cos \theta^\mu > 0) - \sigma(\cos \theta^\mu < 0)}{\sigma(\cos \theta^\mu > 0) + \sigma(\cos \theta^\mu < 0)}, \\ A_C^\mu &= \frac{\sigma(\mu^-) - \sigma(\mu^+)}{\sigma(\mu^-) + \sigma(\mu^+)}, \end{aligned} \quad (18.17)$$

and these are equal. The reason for the equality of these two asymmetries is the  $CP$  invariance of the gauge Lagrangian, even if the  $Z$  has parity violating interactions. Using (18.15) and (18.16) one can calculate the  $A_{FB}^\mu$ , which in general depends on  $s$ . For two different values of  $s$  of interest, it can be shown that:

$$\begin{aligned} A_{FB}^\mu \Big|_{s \ll M_Z^2} &= -\frac{3}{\sqrt{2}} \frac{G_\mu s}{e^2} g_A^2 \frac{1}{1 - \frac{4G_\mu s}{\sqrt{2}e^2} g_V^2}, \\ A_{FB}^\mu \Big|_{s=M_Z^2} &\sim \frac{g_A^2 g_V^2}{(g_A^2 + g_V^2)^2}. \end{aligned} \quad (18.18)$$

In the first case  $M_Z$  drops out as we have made an approximation where  $s \ll M_Z^2$ . In the second case in (18.18), while writing the value for  $\sqrt{s} = M_Z$ , we have used the fact that  $M_Z/\Gamma_Z \gg 1$  and hence the dependency on the precise value of  $M_Z$  drops out. The small width is guaranteed by the weak nature of the NC couplings of the  $Z$  with the fermions. The factor in the denominator of  $\chi$  gives a characteristic resonant shape to the cross-section

for the process  $e^+e^- \rightarrow \mu^+\mu^-$ , the interference term being negative causing the cross-section to reduce below the value expected for the  $\gamma$  exchange alone and to start rising again as  $\sqrt{s}$  approaches  $M_Z$ .

For  $\sqrt{s} \ll M_Z$  the value of  $A$  differs from 1, the value expected in QED, by  $G_\mu g_V^2 s / \sqrt{2} \pi \alpha_{\text{em}}$ . Further the coefficient of the asymmetric, linear term in  $\cos\theta$  is given by the same expression with the replacement of  $g_V^2$  by  $g_A^2$ . Thus it is possible to get information on both  $g_V^2$  and  $g_A^2$  from measurements of  $A$  and  $B$  even with beam energies that are much lower than  $M_Z$ . Since  $G_\mu \sim 10^{-5}/M_p^2$ , the effects can become substantial only when  $s \sim \mathcal{O}(10^4 \text{ GeV}^2)$ . Indeed the first hints of weak NC in this process were obtained in  $e^+e^-$  collisions with  $\sqrt{s} \sim 35 \text{ GeV}$ .

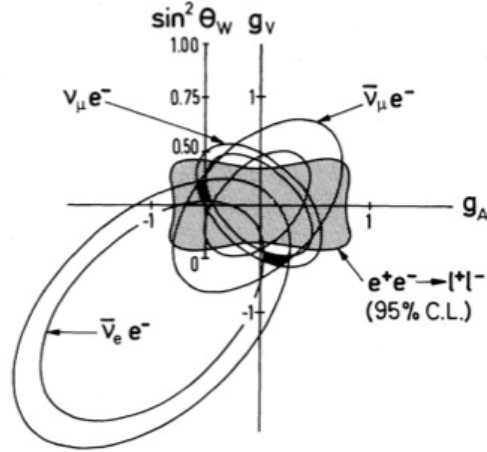
It is worth noting at this point that the calculation of cross-section for quark production via  $\gamma/Z$  exchange proceeds exactly in the same manner, except the expressions will involve  $g_A^q, g_V^q$  in addition to  $g_V^e, g_A^e$  in (18.15) and (18.16). All the observations about  $e^+e^- \rightarrow \gamma/Z \rightarrow \mu^+\mu^-$  then apply for the  $e^+e^- \rightarrow \gamma/Z \rightarrow q\bar{q} \rightarrow \text{hadrons}$  as well.

Note that just like the various cross-sections in Tab. 18.1, the asymmetries of (18.17) and (18.18) too, depend only on one unknown quantity,  $\sin^2\theta_W$  through the vector and axial vector NC couplings of the charged lepton. The above expressions tell us therefore, that a study of the leptonic scattering processes given in the Tab. 18.1 along with the energy dependence of the FB asymmetry and that of the cross-section for the reaction given in (18.15) and (18.16), can provide information about  $\sin^2\theta_W$  much before reaching the beam energies close to  $M_Z$ . If all the measurements of the leptonic cross-sections as well as the asymmetries yielded a unique value of  $\sin\theta_W$ , which is the only free parameter of the model, this can then provide a quantitative validation of the Electroweak model. It is interesting to note that the energy dependence of the cross-section  $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$  can also provide indirect information about  $M_Z$ , much before the energy values close to  $M_Z$  are reached.

Note that production of hadrons by weak NC processes while being very useful for validation of the weak NC due to the large rates possible with nuclear targets, also needed knowledge about the nuclear structure functions to interpret the data. Both the theoretical and experimental understanding of this structure at that time was somewhat rudimentary. Hence the validation of the SM would be much more unambiguous, if one would extract  $\sin^2\theta_W$  using pure leptonic processes alone, the  $\nu$ -charged lepton scattering



and  $e^+e^-$  collisions.

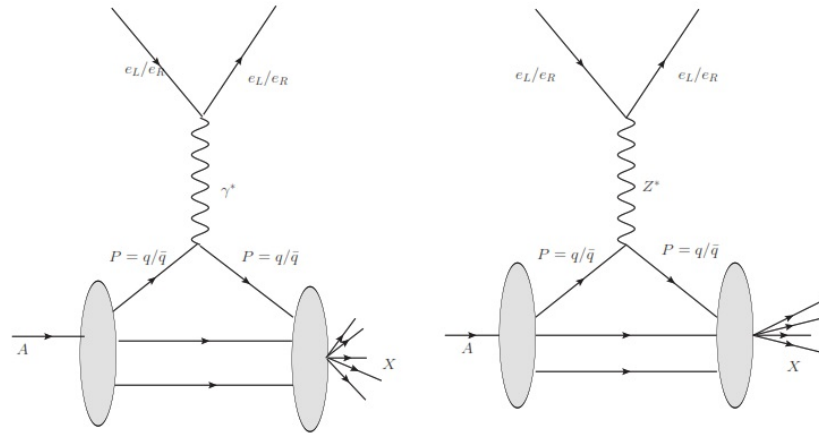


The above figure shows compilation of extractions of  $g_V$ ,  $g_A$  and hence  $\sin^2 \theta_W$  values from pure leptonic processes. These results were among the early quantitative validation of the SM. As explained above the leptonic processes were better suited for a clean and unambiguous extraction of  $\sin^2 \theta_W$ . Further, the CM energies of the early  $\nu$  experiments were limited to  $s < 200 \text{ GeV}^2$ , whereas the  $e^+e^-$  experiments at PETRA at DESY (Hamburg) had  $s \sim 1400 \text{ GeV}^2$ . The  $e^+e^-$  experiments could also probe the NC couplings of the quarks as well, by studying the hadron production along with the  $\mu^+\mu^-$  pair production. Thus the information about the weak NCs at the  $e^+e^-$  colliders was a value addition to the analysis, even though the beam energies were much below than those required to produce an 'on-shell'  $Z$  boson. The left panel shows results on the deviation from the QED expectations of the angular distribution for the  $\mu$  i.e. evidence for both: a nonzero value of  $B$  and value of  $A$  different from 1. It was indeed comparable to the deviation of few percent to be expected at these energies as was argued above. The plot shows comparisons with predictions of the EW model (see (18.18)) for different values  $\sin^2 \theta_W$  showing clear sensitivity to the same. Indeed this as well as measurements of  $\mu$  charge asymmetry defined in the (18.17) for a limited region in the forward hemisphere and the cross-section measurement were used to delineate a region in the  $g_A - g_V$  plane that was allowed by the data at 95% c.l. This is indicated by the grey shaded region in the right panel of the figure. Superimposed on this grey area are also the

regions in the same plane allowed by measurements of  $\bar{\nu}_e e^-$ ,  $\bar{\nu}_\mu e^-$  and  $\nu_\mu e^-$  scattering. We notice from Tab. 18.1 that all the cross-section expressions define different ellipses in the  $g_A - g_V$  plane. The area between two ellipses is the region allowed at 68% c.l. by the measurement of the cross-section for that particular neutrino scattering reaction.

We see from the right panel that if one uses just the elastic  $\nu$ -charged lepton scattering data, there is a twofold ambiguity in the values of  $g_A$  and  $g_V$  that are consistent with the totality of the available data. This is indicated by the two dark black regions. This ambiguity is removed on using the  $e^+e^- \rightarrow l^+l^-$  data. The solution with negative  $g_A$  and positive  $g_V$ , corresponding to the dark region in the upper left corner of the grey shaded square region, is chosen uniquely, after we add determination of  $g_V$ ,  $g_A$  from the  $e^+e^-$  measurements. This dark region in the upper left corner corresponds to  $\sin^2 \theta_W = 0.234 \pm 0.011$ . This was the unique value of  $\sin^2 \theta_W$  consistent with all the 'leptonic' NC measurements mentioned before. One could also use only the  $e^+e^-$  data. Combining all the  $e^+e^- \rightarrow l^+l^-$  measurements with those for  $e^+e^- \rightarrow q^+q^-$ ,  $\sin^2 \theta_W$  was determined to be  $\sin^2 \theta_W = 0.27 \pm 0.08$ . Clearly the two determinations are consistent with each other. These measurements thus conclusively proved existence of the weak NC as predicted by the SM. One could then use the value of  $\sin \theta_W$  so determined, to further make predictions for the  $W$ ,  $Z$  masses as well as their phenomenology.

The weak neutral couplings of the electron can also be probed by studying interference between the  $t$ -channel  $\gamma^*$  and  $Z$  exchange in the Deep Inelastic Scattering (DIS) processes indicated in the figure below, showing weak NC contributions to the deep inelastic scattering with polarised  $e^-$  beams.



This is very similar to the  $e^+e^- \rightarrow l^+l^-$  case. However, in this case one needs to have longitudinally polarized electron beams, to be able to see the effect experimentally. The diagram with  $\gamma^*$  exchange will give a symmetric result for both left and right polarized  $e^-$  but the  $Z$  treats them differently. Recall here that  $g_L^e$  and  $g_R^e$  have different values. Thus there will be a polarization asymmetry in the cross-section. At lower energies and hence smaller values of the invariant mass  $-Q^2$  of the exchanged  $\gamma^*/Z^*$ , it is the interference term between the two diagrams which dominates the size of the observed polarization asymmetry and hence the evidence for parity violation. The interference effect can be shown to be  $\sim G_\mu s$  in this case as well and is linear in  $g_V^e$ . As mentioned before, for the value of  $\sin \theta_W$  realized in nature the vector coupling of the electron is very small. Hence an asymmetry which is linear in this small parameter, provides a more sensitive probe of  $g_V^e$  than the one provided by the asymmetry  $A_{FB}^\mu$  of (18.17). Measurements of this asymmetry also yielded a value of  $\sin^2 \theta_W$  consistent with the determination from the pure leptonic probes.

Finally the best determination of  $\sin^2 \theta_W$  came from high statistics data on  $\nu$ -induced Deep Inelastic Scattering and polarised  $e$ -Deuterium scattering (both not discussed here at all) and the value was:

$$\begin{aligned} \sin^2 \theta_W &= 0.224 \pm 0.015, \rho = 0.0992 \pm 0.017, \\ \sin^2 \theta_W &= 0.229 \pm 0.009 \quad \text{assuming } \rho = 1. \end{aligned} \quad (18.19)$$

In the first case both  $\rho$  and  $\sin^2 \theta_W$  were taken to be unknown and fitted to the data and in the second case  $\rho$  was fixed at 1. Thus  $\rho$  was determined to be  $\sim 1$  as expected in the EW model. Assuming this, around 1981 one could then predict:

$$M_W \simeq 78.15 \pm 1.5 \text{ GeV}, \quad M_Z \simeq 89 \pm 1.3 \text{ GeV}. \quad (18.20)$$

This then sets the goal posts to design experiments which could produce  $W, Z$  directly and study them. In principle, the predictions above receive radiative corrections. We will come to that in the next section.

So the take home message of the above discussion is that the early  $\nu$  experiments as well scattering experiments with polarized electron beams and nuclear targets, along with the  $e^+e^- \rightarrow l^+l^-$  experiments, tested the structure of the NC couplings of the leptons and those of the quarks predicted by the EW model. The experiments conclusively proved that all the measurements were consistent with a unique value of the one undetermined parameter of the model  $\sin^2 \theta_W$ . This then also predicted a narrow range of possible

masses for both the  $W$  and the  $Z$  bosons. Inter alia, these measurements also established  $\rho \simeq 1$ , consistent with the EW prediction again. Thus at this stage, apart from the direct verification of the tree level  $ZWW$  coupling which must exist in this gauge theory, all the other tree level predictions of the model seemed to have been tested.

Given the knowledge of the quark content of the  $p$  available from the DIS experiments, it was also possible to predict the rate of production of these bosons in the process

$$\begin{aligned} p + \bar{p} &\rightarrow W + X \rightarrow l + \nu_l + X , \\ p + \bar{p} &\rightarrow Z + X \rightarrow l^+ + l^- + X . \end{aligned} \tag{18.21}$$

In fact the CERN super proton synchrotron (SPS) was converted into  $Spp\bar{p}S$ , to collide protons on antiprotons, so as to have enough energy to produce the  $W, Z$  in the  $p\bar{p}$  collisions. The observation of the  $W$  and the  $Z$  bosons in the UA-1 and UA-2 experiments, with mass values and production rates which agreed with these predictions, was a very important step in confirming the correctness of the Electroweak model. Later data confirmed the  $V-A$  coupling of the  $W$  bosons to fermions from the angular distribution of the events, even though the original observation had only a handful of these: 6 in UA-1 and 4 for UA-2.

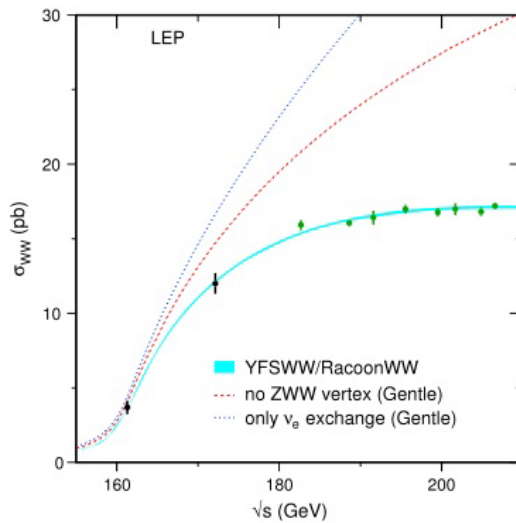
The masses of the  $W$  and the  $Z$  measured in the UA2 experiment, for example, were

$$M_W = 80 + 10 - 6 \text{ GeV} , \quad M_Z = 91.9 \pm 1.3 \text{ (stat)} \pm 1.4 \text{ (syst)} \text{ GeV} . \tag{18.22}$$

The larger errors for  $M_W$  reflect the uncertainties in the measurement of 'missing' transverse momentum due to the  $\nu$  which evades detection. For  $M_Z$ , the first number indicates the statistical error and the second systematic. The use of final state containing leptons allowed for much more accurate determination of the invariant mass in the case of the  $Z$  boson. These masses were certainly consistent with the predictions: see, for example (18.20). One can in principle extract  $\rho$  and  $\sin^2 \theta_W$  from this 'direct' measurement of masses (in particular the accurate measurement of  $M_Z$ ) and compare these with the values obtained from the earlier 'indirect' information from  $\nu$  scattering, for further tests of the SM. This already used the more accurate predictions using energy dependence of the couplings as well EW corrections to the weak processes used to extract  $\sin^2 \theta_W$ . We will discuss this in the context of precision testing of the SM.

## 18.5 Direct Evidence for the $ZWW$ Coupling

Before moving on to the discussion of calculation and validation of loop effects in the precision measurements of the EW observables, we need to discuss the validation of the existence of another tree level coupling of the gauge bosons, the triple gauge boson  $ZW^+W^-$  coupling which is characteristic of the non-abelian nature of the gauge theory. As already discussed, contribution of the  $Z$  exchange diagram is crucial in curing the bad high energy behavior of the  $e^+e^- \rightarrow W^+W^-$  cross-section.  $W^+W^-$  pair production in  $e^+e^-$  collisions was studied at LEP-II where the centre of mass energy was increased from the  $Z$ -pole value of 91 GeV to the two  $W$  threshold of 161 GeV and then finally to 209 GeV. The figure below shows the LEP-II data on energy dependence of the  $W^+W^-$  cross-section, along with the theory prediction.



The data is well described by the solid line which represents the sum of the contribution of the  $\nu_e$  exchange diagram and  $Z/\gamma$  exchange diagrams shown in the left and the central panel of the figure. One sees that the contribution to the cross-section of just the  $\nu_e$  exchange diagram of the left most panel, shown by the blue dashed curve, rises very fast with energy. The cross-section after including contribution of the s-channel  $\gamma$  exchange alone, where the  $ZWW$  coupling is put to zero in the diagram in the central panel of the

figure above, is shown by the red dashed curve. This addition tames the bad high energy behavior to some extent but not completely. Only after adding the  $s$ -channel  $Z$ -exchange diagram does the cross-section have a good high energy behavior, shown by the blue-green solid curve which also describes the data well. Thus we see that the temperate energy dependence of the  $e^+ + e^- \rightarrow W^+W^-$  cross-section shown by the data, is 'direct' proof of the  $ZW^+W^-$  triple gauge boson coupling.

The threshold rise of this cross-section also offers an accurate determination of  $W$  mass and the width:

$$M_W = 80.376 \pm 0.033 \text{ GeV} , \quad \Gamma_W = 2.195 \pm 0.083 \text{ GeV} . \quad (18.23)$$

The same experiment offered a precision measurement of the hadronic decay width of the  $W$  as well. These measurements served later as an input to the precision analysis of the EW observables which we will discuss in the next section.

Note further also that since the energy dependence of the total cross-section is crucially decided by the  $ZW^+W^-$  coupling, it is possible to use the energy dependence and the angular dependence of the process to probe any possible deviations of the  $ZWW$  vertex from the SM structure and value. This process can therefore be successfully used to look for deviations of this coupling from the SM prediction. In view of the important role played by the  $ZWW$  coupling in curing the bad high energy behavior of the  $W$ -pair production cross-section, it is theoretically very important to probe its possible deviations from the SM predictions so as to get indications, if any, of the physics beyond the SM (BSM physics). Measurements of the cross-section and angular distributions of the produced  $W$  at LEP-II, constrained strongly any anomalous  $ZWW$  couplings; i.e. couplings which differ from the SM in either structure or strength.

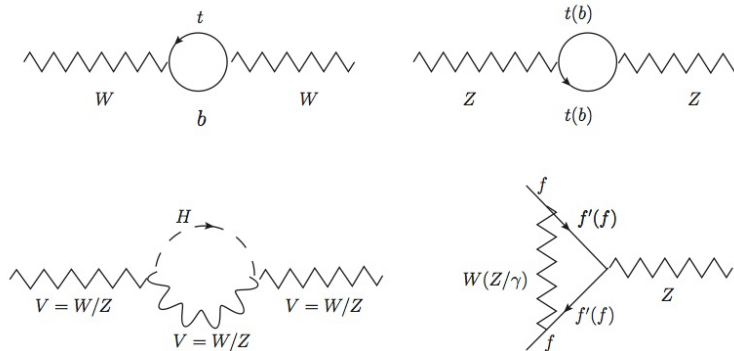
## 18.6 Precision Testing of SM

Thus we see that the various lepton-lepton and lepton-hadron scattering experiments along with the  $p\bar{p}$  experiments helped establish the correctness of Electroweak model predictions at the tree level. These tested the tree level SM predictions for the new NC couplings of the  $Z$  boson with all the known fermions as in terms of the single 'free' parameter of the model. The

prediction of  $SU(2)_L$  symmetry for the structure and strength of the  $ZWW$  vertex was also tested. Last but not the least the experiments also tested the correctness of the tree level predictions for the  $W$  and  $Z$  masses. This indeed established the  $SU(2)_L \times U(1)_Y$  structure of the EW gauge theory. However, even with the somewhat imprecise determined values of  $W, Z$  masses, the need for including the effects of loop corrections, an essential feature of QFT's, on all these tree level predictions was already clear. Since the effect of radiative corrections on the extraction of  $\sin^2 \theta_W$  is different for different processes, it is necessary to correct the experimentally extracted value for these effects, before the  $\sin^2 \theta_W$  extracted from various observables can be compared at high precision.

### 18.6.1 Radiative Corrections and $\rho/\sin^2 \theta_W$

In case of the SM, a QFT with SSB, renormalisability of the theory guarantees that the loop corrections to the tree level relations, will be finite and can be computed order by order in perturbation theory. Precision measurements can then test these corrected relations and hence the correctness of these calculations of loop effects. This can then help establish the renormalisability of the  $SU(2)_L \times U(1)_Y$  gauge theory of EW interactions. Below follows an extremely sketchy discussions of the issues involved.



Some of the one loop diagrams EW corrections to vertices and two point functions in the SM are shown in the above figure. The two diagrams in the top row and the diagram on the left in the lower panel are the ones that need to be considered while calculating the loop corrections to the masses

$M_W, M_Z$ . The diagram on the right in the lower panel is an example of diagrams that give rise to corrections to the  $Zf\bar{f}$  vertex. The dominant corrections come from loops containing quarks of the third generation viz.  $t, b$ . We already notice that corrections to the  $W$  and the  $Z$  mass will be different, since the former involves a  $tb$  loop, whereas the latter involves the  $t\bar{t}, b\bar{b}$  loops. As a result the corrections to  $\sin^2\theta_W$  from these diagrams, for example, will be different for the CC and NC processes. One needs to take into account radiative corrections to the weak processes used to extract  $\sin^2\theta_W$  as well as the energy dependence of the couplings and hence of  $\sin^2\theta_W$ . The latter too is an integral part of QFT. The extraction of  $\sin^2\theta_W$  from weak processes, taking into account all the weak corrections yielded

$$\sin^2\theta_W(M_W) = 0.215 \pm 0.010 . \quad (18.24)$$

In the calculation one now needs to use  $\alpha_{\text{em}}(M_W) = 1/127.49$  instead of the value  $\alpha_{\text{em}} = 1/137.03$ . The expression for  $M_W(M_W)$  then becomes

$$M_W(M_W) = \sqrt{\frac{\pi}{\sqrt{2}G_\mu} \frac{\alpha_{\text{em}}(M_W)}{\sin^2\theta_W(M_W)}} = \frac{38.6}{\sin\theta_W(M_W)} \text{ GeV} . \quad (18.25)$$

This then gives,

$$M_W = 83.5 \pm 2.2 \text{ GeV} , \quad M_Z = 94.2 \pm 1.8 \text{ GeV} . \quad (18.26)$$

Thus loop effects change the predicted values in (18.20) by  $\sim 5\%$ . This sets the scale for the precision with which one needs to measure the masses of  $W$  and  $Z$  to be able to test theory at loop level. The UA-1 and UA-2 measurements were clearly consistent with these predictions within the accuracy of the measurement as well as predictions. These loop corrected predictions for  $M_W$  and  $M_Z$  were used to extract both  $\sin\theta_W(M_W)$  and  $\rho$  just from the measured masses of the  $W$  and  $Z$  in the UA-2 experiment, yielding

$$\sin^2\theta_W = 0.226 \pm 0.014 , \quad \rho = 1.004 \pm 0.052 . \quad (18.27)$$

This value of  $\rho$  is consistent with the expectation of the SM i.e. the Electroweak model where  $W/Z$  masses are generated via SSB. These values are also consistent with the corresponding determinations from the lower energy  $\nu$  experiments (see (18.19)). Agreement of these two independent determinations of  $\rho$  and  $\sin^2\theta_W(M_W)$  from two completely different sets of measurements, already showed consistency of the measurements with theory predictions at loop level.



Diagrams shown above cause  $\rho$  to change from 1, the prediction at tree level, since the corrections are different for  $M_W^2$  and  $M_Z^2$ . In fact, one can write

$$\Delta\rho = \frac{\Sigma_Z(0)}{M_Z^2} - \frac{\Sigma_W(0)}{M_W^2}, \quad (18.28)$$

where  $\Sigma_V$  and ( $V = W/Z$ ) are the one loop corrections to the propagator. As emphasized above these are different for  $W$  and  $Z$  and hence  $\Delta\rho$  is different from 0. At one loop one gets, keeping only the dominant corrections  $\propto M_t^2$ ,

$$\rho_{\text{corr}} = 1 + \Delta\rho \simeq 1 + \frac{3G_\mu M_t^2}{8\pi^2\sqrt{2}}. \quad (18.29)$$

Thus one sees that the relation  $\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} = 1$  gets corrected by loop effects. The corrections are finite as advertised before: a result of the renormalizability of the EW theory. Assuming the (at that time) unknown  $M_t$  to be as large as the largest mass in the theory,  $\sim \mathcal{O}(M_W)$ , one finds corrections to the tree level value of unity of  $\rho$ , to  $\sim$  few parts in 1000. Thus one would need a high precision measurements of  $M_W$  and  $M_Z$  to get a precision value of  $\rho$  which can then be contrasted with above prediction given in (18.29). This can then be used to estimate  $M_t$  and comparing it with the experimentally observed value of the  $t$  quark mass.

In reality, indeed this is what happened. The precision measurements at the  $Z$  pole in  $e^+e^- \rightarrow Z \rightarrow f\bar{f}$ , to be discussed momentarily, indicated a value for the top mass  $M_t \simeq 2M_W$  before the top quark was actually discovered. Agreement of the measured mass of the  $t$  at the Tevatron with this value was then a big success story, testing the SM at loop level. For the much higher value of the mass that the  $t$  quark has in real life compared to the  $M_W$  taken in the numerical estimation above, corrections to  $\rho$  in reality are about 1 part in 100 and hence measurable in precision experiments. For future reference, let me also add here that the corrections to  $M_V^2$ , from the third diagram in the figure above involving the  $VH$  loop, depend on the Higgs mass  $M_h$  only logarithmically.

A detailed discussion of the theoretical significance of all important quadratic dependence of these corrections on  $M_t$ , the logarithmic dependence on  $M_h$  and the non-decoupling nature of the corrections to the  $Zb\bar{b}$  vertex from the  $t\bar{t}$  loop, are beyond the scope of the discussion in these lectures. The former comes from violation of the  $SU(2)_L$  invariance, reflected in the mass difference between the two members of the doublet : the  $t$  and the  $b$ .  $\Delta\rho$  is in fact

proportional to  $M_t^2 - M_b^2$ . The loops involving  $h$  and the  $V$  give contributions to  $\Delta\rho$  which depend on the Higgs mass, but the accidental Custodial Symmetry, guarantees that this dependence will be only logarithmic. This is consistent with the so called Veltman screening theorem. The corrections to the  $Zb\bar{b}$  vertex, originating from the triangle diagram, one of which is shown in figure above, also depend on  $M_t$  quadratically. This quadratic dependence, on the other hand has a different source. It arises from contributions of the longitudinal  $W$  bosons in the loop. In a non-unitary gauge this can be seen as coming from the unphysical Goldstone bosons  $\phi^\pm$ , which are 'eaten up' to become the longitudinal degree of freedom of the  $W$ -boson. This then clearly explains the non-decoupling nature of the correction, coming from the proportionality of  $t\phi^\pm$  coupling  $h_t$  or equivalently  $M_t$ . Even when we do not discuss these issues in detail, suffice it be said that the  $M_t^2$  dependence of the vertex correction is the telltale sign of the SSB via the Higgs mechanism. Since the origins of the  $M_t^2$  dependence, or equivalently the non-decoupling nature of the corrections, are quite different for the  $\Delta\rho$  and  $\delta g_{Z\mu\mu}$  and further only the  $\Delta\rho$  receives contribution from the Higgs, it is quite important to confirm both of these independently. Let us now follow the story of precision measurements and comparison with the precision predictions further.

Note here that these corrections can be calculated only if theory is renormalizable. The renormalizability of a gauge field theory with SSB was proved by 't Hooft. This theory necessarily has a physical scalar, the Higgs boson in the spectrum. As we will see shortly, the precision measurements at the LEP-I of the  $Z$  properties along with weak NC couplings of all the fermions, as well as precision measurements of the properties of the  $W$  at LEP-200, tested these corrections. A test at the loop level of the various relations, could then indicate the need for a finite mass for the Higgs and thus could be an indirect proof for the Higgs! However, we have seen that even with a quadratic dependence of  $\Delta\rho$  on  $M_t$  and the large mass  $M_t$ , the effects are only 1 part in 100, it is clear that with the logarithmic dependence of these corrections on  $M_h$ , this program would require indeed very high precision measurements.

### 18.6.2 Precision Measurements at LEP

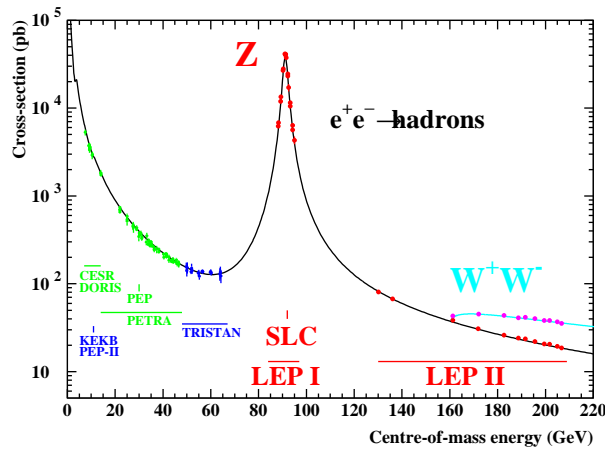
Let us first begin by a discussion of precision measurements of the mass and the coupling of the  $Z$  boson at LEP 1 and the SLC in  $e^+e^- \rightarrow Z \rightarrow f\bar{f}$ . The four LEP experiments studied decays of about 17 Million  $Z$ , whereas

the SLC studied about 600,000  $Z$  decays, but with polarized  $e^+/e^-$  beams. At the end of the day these experiments determined the mass and the width of the  $Z$  boson and also the values of  $\rho$  and effective value of  $\sin^2 \theta_W$ , to a great accuracy using only the leptonic sector. The use of 'effective' implies that radiative corrections have been suitably included while extracting these values.

$$\begin{aligned} M_Z &= 91.1875 \pm 0.0021 \text{ GeV} , & \Gamma_Z &= 2.4952 \pm 0.0023 \text{ GeV} , \\ \rho_l &= 1.0050 \pm 0.0010 , & \sin^2 \theta_{\text{lept}}^{\text{eff}} &= 0.23153 \pm 0.00016 . \end{aligned} \quad (18.30)$$

As already explained these high precision measurements require also high precision calculations, to test the SM at high accuracy. Higher order QCD corrections play a highly important and nontrivial role while using results from the hadronic decays of the  $Z$ . One also requires an excellent understanding of QCD to calculate correctly the observables from quark final states in terms of what the detectors actually observe the jets. This ushered in an era of extremely close and extensive collaboration between experimentalists and theorists resulting in a number of LEP Yellow Reports. These provide the best summary of both the theoretical and experimental issues involved in studies at LEP.

The figure below shows a compilation of the cross-section for the process  $e^+ + e^- \rightarrow \text{hadrons}$ , spanning the entire energy range from PEP/PETRA to LEP II. Solid line is theory prediction, including the electromagnetic and the QCD radiative corrections.



Recall the expression for the cross-section for  $e^+e^- \rightarrow \mu^+\mu^-$  given in (18.16).

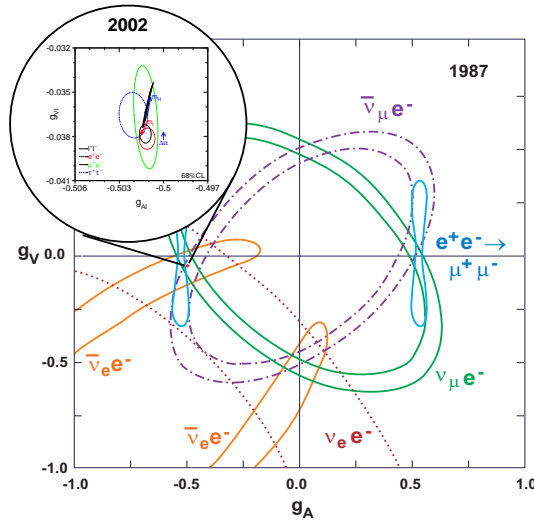
The initial fall off of the cross-section reflects the  $1/s$  dependence of the first  $\gamma$  exchange diagram. One can then see the onset of the rise in the cross-section due to interference between the  $\gamma$  and  $Z$  exchange contributions.

Recall that it is these interference terms, at energies quite far away from the  $Z$  resonance, that had allowed the first glimpse of effects of weak NC in the process  $e^+e^- \rightarrow \mu^+\mu^-$ . Thus we see that the  $Z$  resonance makes its presence felt much before the resonant energy is reached, by just the shape of the cross-section curve. This line shape of the  $Z$  resonance depends on  $\Gamma_Z$ ,  $M_Z$ , partial decay width  $\Gamma(Z \rightarrow f\bar{f})$  and through them on  $g_V$ ,  $g_A$  of the electron and the fermions in the final state being considered.

The extremely accurate measurements of  $M_Z$ ,  $\Gamma_Z$  mentioned above, were extracted by fitting the shape of this curve near resonance, taking into account effects such as the initial state radiation etc. This precision study of the line shape of  $Z$  was made possible by the unprecedented energy resolution of the collider LEP-I. The thin solid line is then the theoretical prediction for the cross-section including the QED and QCD radiative correction. The asymmetric shape of the curve near the resonance is the effect of the initial state radiation. The agreement between the data and theory needs no comment.

Recall now the discussion in Sec. 18.4 and (18.16) – (18.18). One can extend constructions of these asymmetries of (18.16) – (18.18), for all the fermionic final states accessible in the  $Z$  decay, the leptons  $e$ ,  $\mu$ ,  $\tau$  and the quarks  $b$ ,  $c$ . Looking at the expressions in (18.16) – (18.18) one can see that a precision measurement of these asymmetries as well as partial widths, lead to an accurate determination of  $g_V^f$  and  $g_A^f$ . The  $Z$ -decay data from SLC, which employed linearly polarized  $e^-/e^+$  beams, allowed for constructing polarisation asymmetries just like the forward-backward asymmetry of (18.18). This too is a measure of parity violation, with the additional advantage that it involves  $g_V$  linearly instead of the quadratic dependence in (18.18). This linear dependence is similar to the case of polarization asymmetries in case of polarized electron-Deuterium scattering mentioned before.

Recall also that for the value of  $\sin^2 \theta_W$  of (18.19) which is rather close to 0.25, the vector coupling of the electron involving  $(4 \sin^2 \theta_W - 1)$  is very small. Hence this linear dependence of the asymmetries on  $g_V$  allowed the experiments at the SLC to reach a competitive accuracy for the extraction of  $g_A$  and  $g_V$  with the much smaller luminosity and hence smaller number of the  $Z$  decays (600,000 versus 17 million at LEP) available there.



The figure above shows values of  $g_V^e$  and  $g_A^e$  obtained using the LEP-I data, juxtaposed with the data from elastic  $\nu$  scattering from 1987. Compare the size of the region in the  $g_V - g_A$  plane selected by all the measurements (shown in an inset at the left). We see that at the  $Z$  pole the weak NC couplings of the  $Z$  with the fermions, were tested to about one part in 1000.

It goes without saying that with such precision in measurements, if one were to repeat the earlier exercise of extracting the value  $\sin^2 \theta_W$ ,  $\rho$  from them, such as given in (18.30), one has to use theoretical predictions which include all the relevant higher order corrections. This was already discussed in Sec. 18.6.1. Since these corrections have a dependency on the masses of the particles like the  $W$ ,  $t$  and the Higgs, if the measurements are precise enough then they can be sensitive to these masses. The precision measurements of EW observables then indicate 'indirectly', in the framework of the SM, the values of the masses of these particles preferred by the precision EW data. A comparison of these masses determined 'indirectly', with the ones measured directly, can then be a powerful precision test of the SM.

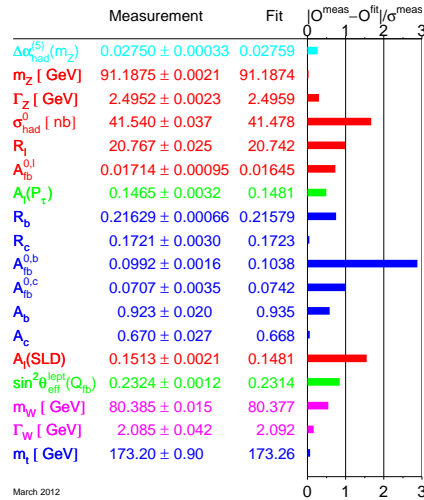
### 18.6.3 Precision Testing and Indirect Bounds

The EW part of the SM has following free parameters:  $g_1$ ,  $g_2$ ,  $v$  and  $\lambda$ . At tree level all the couplings of the gauge bosons to fermions as well as to each other and their masses are completely given in terms of the first three

parameters in this list. With the very precise knowledge of  $M_Z$  provided by the LEP-I, it made sense to trade the  $g_1$ ,  $g_2$  and  $v$  for  $M_Z$ ,  $\alpha_{\text{em}}$  and  $G_\mu$ , to express all the EW observables as functions of these three chosen quantities.

A large number of EW observables have been measured very accurately, beginning from the total width of  $Z$  boson, the various forward-backward and polarization asymmetries on the  $Z$ -pole, masses  $M_W$ ,  $M_t$ , polarized  $e$ -Deuterium scattering, atomic parity violation etc. All these observables depend on  $G_\mu$ ,  $M_Z$  and  $\alpha_{\text{em}}$  through their dependencies on  $g_A^f$ ,  $g_V^f$ ,  $M_V$  as well as on  $\alpha_s$  and  $M_t$ ,  $M_h$  through the higher order QCD and EW corrections.

Precision calculation for all these EW observables, including the 1 loop EW radiative corrections in the framework of the SM, are available. The idea is to make then a fit to the measured values of the EW observables and test the SM predictions. In these fits, one keeps  $M_t$ ,  $M_W$  and  $M_h$  as free parameters. As already noted the radiative corrections depend on  $M_t$  quadratically and  $M_h$  logarithmically. Then compare the  $M_W$ ,  $M_t$  values so obtained with experimentally determined values of the same, thus providing a test of the SM. Afterwards one can perform the exercise by varying the Higgs mass, find the value of  $M_h$  that minimizes the  $\chi^2$  and then find the limits on the Higgs mass for which the data will be consistent with the SM predictions.



The figure above shows the result of such an exercise. The figure lists the measured values of a variety of EW observables, most of which we have discussed. The various  $R$ -ratios:  $R_b$ ,  $R_c$ ,  $R_l$  etc. are a measure of the relative production of the various final states and hence of the partial decay width of the  $Z$  into them.  $A_l(P_\tau)$  is the polarization asymmetry for the  $\tau$ 's produced in  $e^+e^- \rightarrow Z \rightarrow \tau^+\tau^-$  on the  $Z$ -pole. The second column shows the result of the SM fit for the observable and the third column the pull which is the difference between the measurement and the fit value normalized by the error of the measurement. The pull is less than three for all the observables and above 2 for only one of the measurements  $A_{FB}^b$ .

This particular fit is the last one before the discovery of the Higgs at the LHC, using the most accurate measurement of  $M_W$  from the Tevatron, which has an error of 0.15 GeV, again a 'one per mille' measurement. The  $\chi^2$  of this fit is not very small, mainly due to the discrepancy between the best fit values and measured values for  $A_b$  from LEP as well as at the SLC. Hence before the 'direct' discovery of the Higgs there were a few physicists who used to be a little uncomfortable about the goodness of the fit and accepting this as 'the proof' for the correctness of the SM at loop level.

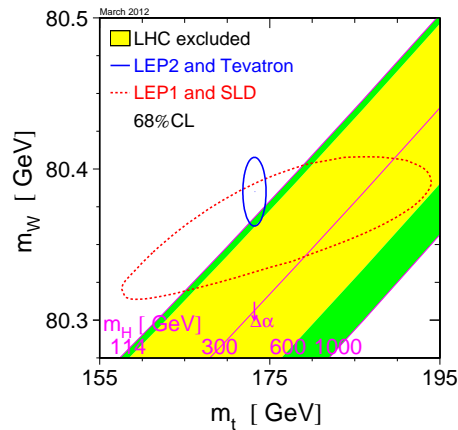
Note the values in the last two rows. The measured values and the best fit values of  $M_W$ ,  $M_t$  agree with each other to a great precision and the pull is rather small, providing thus a stringent test of the SM at loop level. This is the agreement between the  $M_t$  predicted 'indirectly' from the LEP EW precision measurements and the 'direct' measurement from the Tevatron, that was alluded to before a few times. In fact this spectacular agreement was the QFD (Quantum Flavor Dynamics) equivalent of testing the  $(g-2)_\mu$  prediction with the measurement in QED. The important role played by renormalizability and loop corrections in this context can be understood by doing a small numerical exercise of predicting  $M_W$  from the very accurately measured values  $\alpha_{em} = 1/137.0359895(61)$ ,  $G_\mu = 1.16637(1) \times 10^{-5} \text{ GeV}^{-2}$ ,  $M_Z = 91.1875 \pm 0.0021 \text{ GeV}$  and the tree level relations given by the SM among these quantities and  $M_W$ . By using the tree level relation  $M_Z = \frac{M_W}{\cos \theta_W}$  we can write:

$$\frac{G_\mu}{\sqrt{2}} = \frac{g_2^2}{8M_W^2} = \frac{\pi\alpha_{em}}{2M_W^2(1 - M_W^2/M_Z^2)}. \quad (18.31)$$

This gives,  $M_W^{tree} = 80.939 \text{ GeV}$ . Compare this now with the value of  $M_W$  given in the second column of the table above,  $M_W^{expt} = 80.385 \pm 0.015 \text{ GeV}$ .

Of course, this points out the need for calculating loop corrections to the tree level relations. Renormalizability guarantees that all the corrections are finite and can be computed. Hence the value of  $M_W$  obtained 'indirectly' from the fits using theoretical predictions which include these loop corrections, then famously agrees with the 'direct' measurement as shown in the table above. Agreement with the SM would have been impossible unless the predicted values included higher order corrections calculated in perturbation theory.

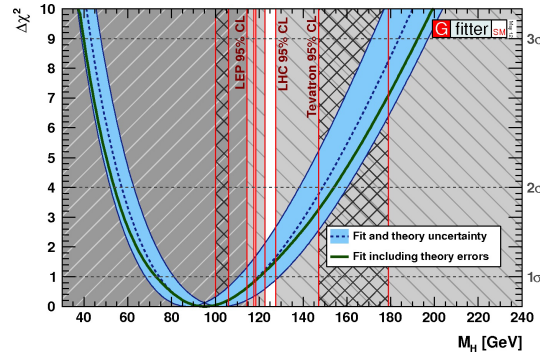
The fit values and the pull for  $M_t$ ,  $M_W$  depends on the value of  $M_h$ , albeit very weakly, due to the logarithmic dependence on  $M_h$  of the EW corrections to  $M_W$ ,  $M_Z$  etc. Some of these effects can be seen from the two figures below made before the Higgs discovery (March 2012).



This plot shows the dependence of the fit values for  $M_W$ ,  $M_t$  for different values of  $M_h$ . The long lopsided ellipse used the EW observables measured at LEP-I and the SLC, to determine allowed regions in the  $M_t$ - $M_W$  plane at 95% c.l. Using the  $M_W$  measurements at the LEP-II/Tevatron as input, one now obtains the small blue ellipse which is consistent with the precision measurements. The dark green (grey) region and the large red ellipse show that with results from LEP-I alone, the measurements were not sensitive to  $M_h$  at all. On the other hand, the highly accurate LEP-II/Tevatron measurements of  $M_W$  and the Tevatron measurement of  $M_t$  is consistent with somewhat small values of the Higgs mass at the left most boundary of the green(grey) region. This was also consistent with the exclusion (from



direct searches at the LHC) of a SM Higgs over a very large range as indicated by the  $M_h$  values labeling the inclined lines in the region shaded in yellow (a shade of lighter gray).



The above plot shows the same information in a different format, where we show a plot of  $\Delta\chi^2$  as a function of  $M_h$ . In fact the fact that this minimum of  $\Delta\chi^2$  occurs at a nonzero, finite mass  $M_h$  is already an indication of the 'existence' of the Higgs and hence a feather in the cap of the SM. The dotted and solid black lines are the best fit with and without including the theory errors. The region shaded in light blue (grey) indicates effect of the theoretical uncertainties as well as uncertainties in the EW fit. In the absence of any information from 'direct' searches for the Higgs, the indirect constraints will allow a region around the minimum of  $\chi^2$  ( $M_h \simeq 90 - 100$  GeV) upto  $M_h$  values where  $\Delta\chi^2$  is 9: the  $3\sigma$  value. Remaining values of  $M_h$  will be disfavored by this 'indirect' search. The  $\Delta\chi^2 \leq 9$  corresponds to an allowed mass range  $40 - 45 \sim M_h \sim 180 - 200$  GeV at  $3\sigma$ . However a lot of this 'allowed' region is ruled out from direct searches at the LEP, at the Tevatron and at the LHC. These bounds are indicated by the vertical red lines in this figure. The region ruled out by LEP is indicated by the dark grey region hatched with slanted lines. The region ruled out by the hadronic collider Tevatron is indicated by the cross-hatched region. The above mentioned red lines mark the edges of these regions giving us the pre-LHC exclusion. The region excluded by the LHC in March 2012 is indicated by light grey region marked by lines slanted in a direction opposite to the LEP exclusion region.

As one can see from this figure, before the LHC direct search constraints, the allowed mass range for the Higgs was  $115 \leq M_h \sim 150 - 160$  and

$180 \sim M_h \sim 200$  GeV. The LHC experiments ruled out existence of an SM Higgs in a major part of this range. As a result in March 2012, the mass value allowed for a SM Higgs by a combination of the EW precision measurements and 'direct' collider constraints was as indicated by the small white slit around 125 GeV. Failure to find a Higgs in this small 'allowed' mass range would then have meant the death for the SM. Indeed a new boson was found with properties very similar to a SM Higgs in precisely this mass range. This discussion should make it very clear to us that the value of the mass of the observed Higgs boson itself tested the SM at loop level to a very great accuracy.

In fact it won't be out of place to recapitulate at this point how the SM was validated and tested at various levels by discovery of new particles whose masses were predicted: either in terms of a free parameter of the model which could be determined from experiments OR 'indirectly' by comparing loop effects on physical observables with their precision measurement.

- Observation of suppression of FCNC implied that the quarks must come in isospin doublets. Thus charm was predicted since the existence of the  $s$  quark was known and top was predicted to be present once the  $b$  was found. Further, the very demand of cancellation of anomalies so as avoid these spoiling the renormalizability, implied existence of third generation of quarks and leptons once the  $\tau$  was found.
- One could get indirect information on  $M_c, M_t$  from flavor changing NC processes induced by loops. Agreement of this 'indirect' information with 'direct' measurements 'proved' the correctness of description of EW interactions in terms of a gauge theory.
- CP violation in meson systems could be explained in terms of the SM parameters and measured CKM mixing in quark sector only if three generations of quarks exist.
- $M_W, M_Z$  was predicted in terms of  $\sin\theta_W$  and direct observation of the  $W, Z$  at the predicted mass tested the particle content and tree level coupling of the matter fermions with the gauge bosons  $W, Z$ .
- Study of energy dependence of the  $e^+e^- \rightarrow W^+W^-$  process gave direct evidence for the tree level  $ZWW$  coupling and also for the role played by this vertex in taming the bad high energy behavior of the cross-section.

- Further, Tevatron found evidence for 'direct' production of the top quark at the mass  $M_t$  which was in agreement with the value obtained 'indirectly' from precision measurement of  $M_W$ ,  $M_Z$ , considering effect of radiative corrections to these masses.
- Last but not the least the existence of a minimum of  $\Delta\chi^2$  at a finite nonzero mass for the SM fits to the EW precision measurements, gave an 'indirect' proof of the existence of the Higgs. Before the 'direct' discovery of the Higgs this was also an 'indirect' probe of the couplings of the Higgs with gauge bosons and the  $t$  quarks. Further, the same fits gave an 'indirect' determination of  $M_h$  which now agrees completely with the measured mass of the observed Higgs.

Now we can turn once again to the discussion of the above plots. As was already indicated by the first plot, the 'directly' measured value of the Higgs mass

$$M_h = 125.09 \pm 0.24 \text{ GeV} \quad (18.32)$$

is right in the 'allowed' white slit and indeed confirms the SM at loop level most spectacularly. At this point, it is worth noting that if we improve upon the accuracy of measurements of  $M_t$ ,  $M_w$  and  $M_h$  we can indeed hope to look for effects by loops of heavy particles which are not present in the SM but are expected to exist in various extensions of the SM, which are in turn postulated to address various shortcomings of the SM!

As already mentioned, the Higgs mass range allowed by the EW precision measurements can change when one goes away from the SM. In fact before the 'direct' discovery of the Higgs, a lot of effort had gone on, in constructing models which would allow one to avoid these constraints, should experiments reveal a Higgs boson not consistent with the bounds from the EW precision measurements.

Of course, not only that many of these are not required, but some are now even ruled out, by the observation of the light state. An example of one such model is the SM with a fourth sequential generation of fermions, leptons and quarks. Since in the SM there is no guiding principle for total number of generations of fermions, except that they should be the same for quarks and leptons, this in principle is the simplest extension of the SM by addition of more matter particles to it. Observation of the low mass  $\sim 125$  GeV scalar ruled out this extension very conclusively.

## 18.7 Mass of the Higgs

As we saw above the EW precision measurements did put 'indirect' bounds on the Higgs mass. However, theoretically there is no information on the mass of the Higgs in the SM, as it is determined by  $\lambda$  an arbitrary parameter. Recall  $M_h$  and  $\lambda$  are related by

$$M_h^2 = 2\lambda v^2 . \quad (18.33)$$

The observed mass of the Higgs determines the self-coupling  $\lambda$ :

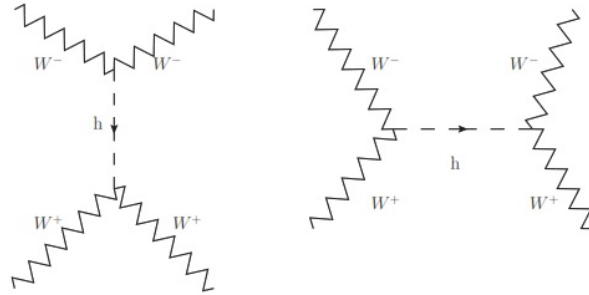
$$\lambda = 0.5M_h^2/v^2 \simeq 0.13 . \quad (18.34)$$

This is the last free parameter of the SM that needed to be determined. Thus the only part of the scalar potential now that needs to be experimentally verified 'directly' is the triple Higgs and the quartic Higgs couplings. Now that one 'knows' the value of  $\lambda$  one can assess the possibilities of measuring it at current and future colliders. One might ask the question whether this is the only nontrivial information about the SM that we can extract from the observed value of the mass of the Higgs. Asked differently, can one use this observed value of  $M_h$  to infer something about the SM as well as the physics beyond the SM, the BSM. Since in these lectures we restrict ourselves to the SM, I will only talk about the possible implication of the observed Higgs mass for the SM itself.

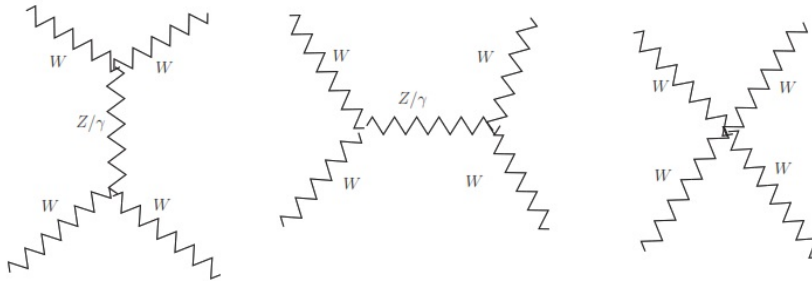
While the SM has no 'prediction' for  $M_h$ , requirement of theoretical consistencies imply bounds on the same. These theoretical limits on the mass of the Higgs boson come from demanding good high energy behavior of scattering amplitudes in the  $SU(2)_L \times U(1)_Y$  gauge theory and from the quantum corrections that the self-coupling  $\lambda$  receives. These limits are thus essentially an artifact of the quantum field theoretical description. Let us discuss this one by one.

### 18.7.1 Unitarity Bound

Various contributing diagrams for the high energy behavior of the scattering amplitude  $W^+W^- \rightarrow W^+W^-$  are shown below. The first panel shows diagrams involving  $h$  bosons contributing to  $W + W \rightarrow WW$  scattering.



The  $s$ -channel diagram will of course contribute only for  $W^+W^- \rightarrow W^+W^-$  scattering. The second panel shows the all the diagrams which involve exchange of the gauge bosons  $Z$  and  $\gamma$  as well as the one involving pure gauge vertex.



Each of these diagrams gives a contribution which grows as  $s^\alpha$  with  $\alpha = 1, 2$ , where  $s$  is the centre of mass energy of the  $WW$ . This divergence appears in the scattering of longitudinal  $W$ 's. However in the SM all the divergent terms in the  $WW \rightarrow WW$  amplitude cancel among each other.

The contribution of the  $h$  exchange diagrams as well as that from the diagrams with pure gauge vertices play an essential role in this cancellation as mentioned before. The cancellation of the power divergences is independent of the Higgs mass and thus the requirement of non-divergent behavior does not single out any scale. Among the non-divergent part of the amplitude  $\mathcal{A}(WW \rightarrow WW)$ , left over after all this cancellations, the contributions of the Higgs exchange diagrams dominate and are dependent on the Higgs mass. The non-divergent part of this invariant amplitude can be written as

$$\mathcal{A}(W_L^+W_L^- \rightarrow W_L^+W_L^-) = -\sqrt{2}G_\mu M_h^2 \left( \frac{s}{s - M_h^2} + \frac{t}{t - M_h^2} \right). \quad (18.35)$$

From a partial wave analysis of this amplitude one can show that this amplitude will violate tree level unitarity if

$$M_h > \left( \frac{8\pi\sqrt{2}}{3G_\mu} \right)^{1/2} \sim 1000 \text{ GeV} . \quad (18.36)$$

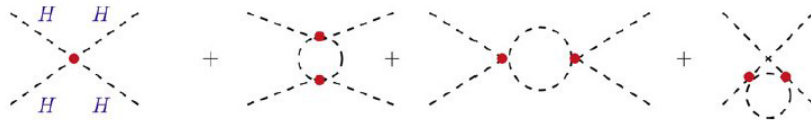
Thus, the theory will be strongly interacting if  $M_h$  were to exceed this value. As things stand, the observed value of  $M_h$  implies  $\lambda \simeq 0.13$ , far from the strongly interacting region and also safe from any unitarity violation. Thus the observed mass of the Higgs boson satisfies the unitarity bound.

### 18.7.2 Triviality Bound

Effect of loop corrections to the self-coupling  $\lambda$  in a scalar field theory, in the presence of a high scale and additional interactions with gauge bosons and matter, with an aim to examine whether one could constrain the scalar mass and other high scale masses from pure theoretical considerations. Triviality bound results from considering loop corrections to the scalar potential. One demands that the quartic coupling  $\lambda$  in the Higgs potential reproduced below,

$$V_h = \lambda v h^3 + \frac{\lambda}{4} h^4 , \quad (18.37)$$

remains perturbative as well as positive at all energy scales under loop corrections. The corrections come from two sets of diagrams shown somewhat schematically in figures below. The first diagram shows loop corrections to the quartic coupling  $\lambda$  from the Higgs sector itself.



The second diagram shows contributions to the running of  $\lambda$  from fermion and gauge loops. So the diagrams shown in the top panel are applicable to any scalar with quartic self-interaction. The ones in the lower panel are specific to a gauge theory.



The *triviality bound* comes from demanding that  $\lambda$  should always remain perturbative. To understand the origin of this bound let us consider the case of large  $M_h$ . Since  $M_h^2 = \lambda v^2$ , at large  $m_h$  and hence large  $\lambda$ , loop corrections are dominated by the  $h$ -loops shown in the first panel above. A straightforward evaluation of this gives us

$$\frac{d\lambda(Q^2)}{d\log Q^2} = \frac{3}{4\pi} \lambda^2(Q^2) . \quad (18.38)$$

Solving this, one gets

$$\lambda(Q^2) = \frac{\lambda(v^2)}{[1 - \frac{3}{4\pi^2} \lambda(v^2) \log(\frac{Q^2}{v^2})]} . \quad (18.39)$$

A look at (18.39) shows us that at large  $Q^2 \gg v^2$ ,  $\lambda(Q^2)$  can develop a pole, the so called Landau pole, at some high scale  $Q$  depending on the value  $\lambda$  at the EW scale  $v$ . If we demand that  $\lambda$  remains always in perturbative regime, then the only solution would be  $\lambda = 0$ . This would then mean that the theory will be trivial. That of course does not make for a sensible theory. Thus the starting value of  $\lambda(v)$  and hence  $M_h$  is not allowed by these considerations.

One can understand this in yet another way. If we demand that the scale at which  $\lambda$  blows up is above a given scale  $\Lambda$ , then using (18.39) we find that for a given value of  $M_h$  and hence  $\lambda(v)$ , the scale at which the Landau pole lies will be given by

$$\Lambda_C = v \exp\left(\frac{2\pi^2}{3\lambda}\right) = v \exp\left(\frac{4\pi^2 v^2}{3M_h^2}\right) . \quad (18.40)$$

Thus, for example, using  $\Lambda_C = \Lambda = 10^{16}$  GeV, we will find  $M_h \sim 200$  GeV.

This bound is called the triviality bound. In simple terms it means that the value of  $\lambda$  at the EW scale (and hence the mass  $M_h$ ) should be small enough so that  $\lambda(Q^2)$  does not develop a pole up to a scale  $Q = \Lambda_C$ . Hence, if  $M_h$  were found to have a mass larger than the triviality bound, it would have meant existence of new physics below the scale  $\Lambda_C$ . This thus tells us that

just the mass of the  $h$  can give us an indication about the scale at which SM must be complemented by additional new physics. The mass of the Higgs being only 125.09 GeV this is rather an academic discussion as this small value of the coupling  $\lambda$  at the EW scale, implies that the loop effects will not be driving the self-coupling  $\lambda$  toward the Landau pole at an energy scale of interest. There are other issues that we need to address given that the observed mass is so small. But we will not discuss them here.

### 18.7.3 Stability Bound

When  $M_h$  is small and  $\lambda$  is not large, the fermion/gauge boson loops are important. Even more important is that the fermions loops come with a negative sign. This means that if the fermion mass is large enough the loop corrections may drive  $\lambda$  negative at some scale, unless the starting value of  $\lambda(v)$  is large enough. These considerations will imply a lower bound for  $\lambda(v)$  and hence for  $M_h$ . This limit on  $M_h$  is called *the vacuum stability bound*. Now one works in the limit of small  $\lambda$ , opposite to the one used when considering the triviality bound.

Hence the contribution of the  $h$ -loops shown in the first panel above can be neglected, the equation for energy dependence of  $\lambda$  now can be written as:

$$\begin{aligned} \frac{d\lambda(Q^2)}{d\log(Q^2)} \simeq & \frac{1}{16\pi^2} [12\lambda^2 + 6\lambda f_t^2 - 3f_t^4 - \frac{3}{2}\lambda(3g_2^2 + g_1^2) + \\ & + \frac{3}{16}(2g_2^4 + (g_2^2 + g_1^2)^2)] . \end{aligned} \quad (18.41)$$

Here

$$f_t = \frac{\sqrt{2}M_t}{v} \quad (18.42)$$

is the Yukawa coupling for the top. Since

$$M_t \sim 173\text{GeV} , \quad v \simeq 246\text{GeV} , \quad (18.43)$$

one can see that the Yukawa coupling is  $\simeq 1$ . Thus it will dominate the scale dependence of  $\lambda$ . At small  $M_h$  and hence small  $\lambda(v)$ ,  $\lambda$  can turn negative at some value of  $Q$ .

Recall the Higgs potential. A negative value of  $\lambda$  will mean an unbounded potential and clearly the vacuum will be unstable. The condition for non-



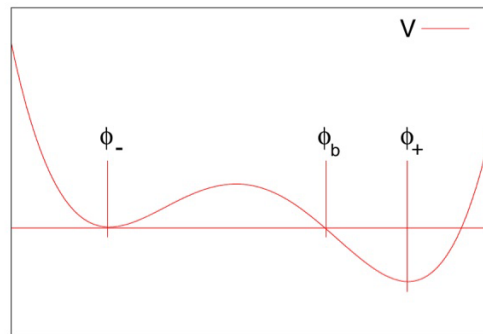
negativity of  $\lambda$  and hence vacuum stability, is

$$M_h^2 > \frac{v^2}{8\pi^2} \log(Q^2/v^2) \left[ 12m_t^2/v^4 - \frac{3}{16}(2g_2^4 + (g_2^2 + g_1^2)^2) \right]. \quad (18.44)$$

Again, depending upon the scale up to which we demand the potential to be positive definite, we find that the starting value  $\lambda(v)$  (and hence  $M_h$ ) has to be above a critical value dependent on the scale. If we demand that the  $\lambda(Q)$  is positive up to  $\Lambda_C$  we then get a lower bound on  $M_h$ . For example choosing,  $\Lambda_C = 10^3$  GeV we get  $M_h \sim 70$  GeV. This bound is called the stability bound.

In the above analysis we have demanded that  $\lambda(\Lambda)$  does not become negative so that the potential is stable. This is the condition for absolute stability of vacuum. However, Planck scale dynamics might stabilise the vacuum for  $|\Phi \gg v$  and we might be living in a metastable vacuum which has a life time bigger than that of the Universe.

The cartoon shown below indicates such a situation.

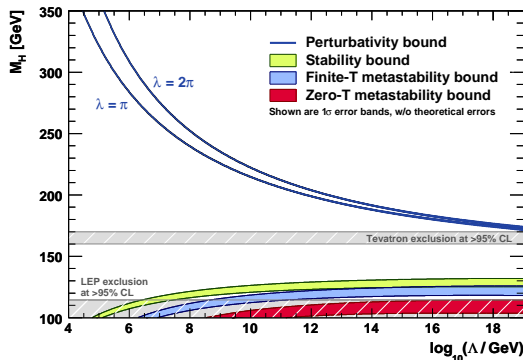


One can then obtain lower bounds on the Higgs mass demanding that vacuum is metastable with a lifetime bigger than the lifetime of the Universe. Clearly evaluation of these bounds cannot be presented in the simplistic analysis given here.

A complete analysis in fact gives the vacuum stability bounds on the Higgs mass taking into account the effect of renormalization group evolution (RGE) as well as that of metastability of the vacuum.

The figure below shows the stability bounds on  $M_h$  as a function of the scale,

indicated by the pale yellow green area, as a function of scale at which the instability sets in.

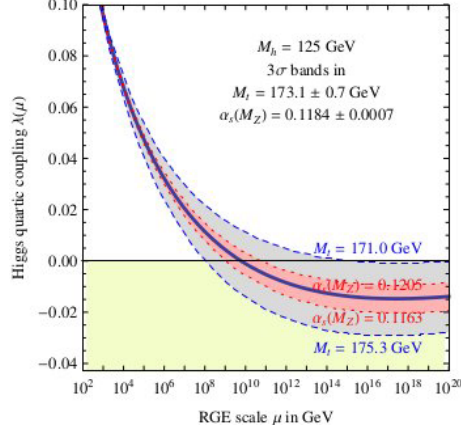


Bounds are shown for absolute stability as well as metastability. The spread is due to the theoretical uncertainties, major ones being the top mass uncertainty and the missing higher order contributions to the equations. RGE takes into account not just the one loop corrections, but also includes the resummation of leading logarithmic corrections.

As one can see even from the simple minded analysis presented here, the bound depends critically on the value of  $f_t$  and hence on  $M_t$ . If one overlays the bounds on the Higgs mass obtained 'indirectly' from the EW precision analysis as well as the LEP/Teavtron/LHC searches then we realize that the thin white silver which was still allowed by March 2013 corresponds to the boundary of the pale yellow-green region indicating the stability bound. Due to the finite width of these bands caused by various uncertainties mentioned above, the observed mass of the Higgs  $M_h$  may or may not be consistent with the hypothesis that the SM remains consistent all the way to Planck scale. Given that everything depends logarithmically on different scales and with the high accuracy of the experimental measurement of  $M_h$ , the need to do the evolution of  $\lambda$  taking into account higher order effects is thus clear.

In fact the need for more accurate calculation was already apparent, even before the Higgs discovery, with the rather low values of  $M_h$  indicated by the 'indirect' limits. The need for accuracy in the theoretical prediction of stability bound is thus very apparent. In May 2012, with the discovery of the Higgs imminent, an NNLO analysis of the problem became available, which reduced the theoretical error on the bounds coming from the unknown higher order corrections to  $\sim 1$  GeV.

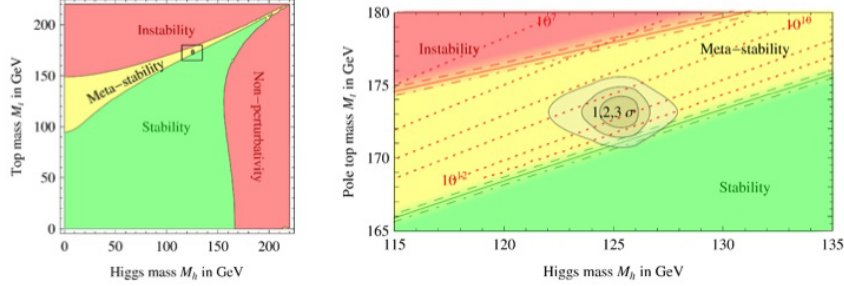
However, there still remains a sizable error due to the errors in experimentally determined parameters  $M_t$ ,  $\alpha_s$ . The figure below shows behavior of  $\lambda(\mu)$  as a function of the energy scale  $\mu$ .



One now sees clearly that the scale at which  $\lambda$  becomes zero and hence the vacuum unstable, depends critically on  $M_t$  and the strong coupling  $\alpha_s$ . For example, for the central value of  $M_t$  used,  $\mu$  value at which  $\lambda$  becomes zero changes by at least an order of magnitude as  $\alpha_s$  is varied within errors. The dependence on  $M_t$  is even stronger. We will comment later on the range of  $M_t$  used in this analysis. According to this analysis the absolute stability of the vacuum up to Planck scale  $M_{pl}$  is guaranteed for,

$$M_h \text{ [GeV]} > 129.4 + 1.4 \left( \frac{M_t \text{ [GeV]} - 173.1}{0.7} \right) - 0.5 \left( \frac{\alpha_s(M_Z) - 0.1184}{0.0007} \right) \pm 1.0 \text{ th.} \quad (18.45)$$

In this analysis the error on pole mass of the top was taken to be  $\Delta m_t = \pm 0.7 \text{ GeV}$ . Taking into account the errors, (18.45) then means that for  $m_h < 126 \text{ GeV}$ , vacuum stability of the SM all the way to Planck Scale is excluded at 98% c.l. Clearly, this value is far too close to the observed value of  $125.09 \pm 0.24 \text{ GeV}$  to require careful considerations of various issues before we draw conclusions about the validity of the SM at high scale. For the measured value of the Higgs mass, the exact scale where  $\lambda$  crosses zero, though not  $M_{pl}$  seems close to it and depends entirely on the exact value of  $M_t$  and  $M_h$ . Indeed these considerations may be relevant for consideration of BSM or models of inflation etc.



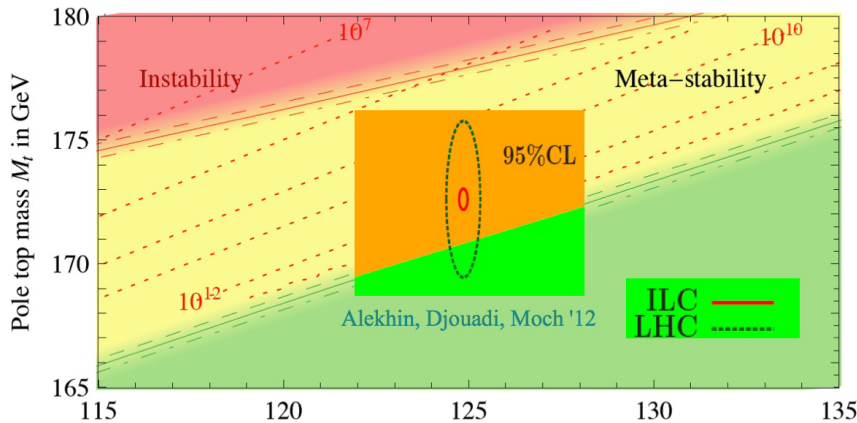
The same can be seen clearly from the figure above. This shows the results of this NNLO analysis of the region in  $M_h - M_t$  plane from the vacuum stability considerations. The left panel shows the regions in the  $M_t - M_h$  plane where the vacuum is absolutely stable, metastable and unstable. To understand the role and size of various 'experimental' uncertainties the right panel shows a zoom in of the region around the experimentally determined  $M_h - M_t$  values. The grey areas show allowed regions at 1,2 and 3  $\sigma$ . The three curves on the boundary of two regions correspond to three values of  $\alpha_s$ . Superimposed on it are the contours of constant value of the high scale where the instability occurs. We see that the experimentally determined values lie right on the boundary of the stable/metastable region. The answer to the question as to whether or not, the experimentally determined value of  $M_h$  is consistent with SM vacuum being (meta)stable all the way to Planck scale, very much depends on  $M_t$  values.

Let us discuss this issue in a little more detail. The stability bounds used errors on  $m_t$  as measured at the hadronic colliders the Tevatron and the LHC. This is the so called Monte Carlo or kinematic mass, which is a parameter in the Monte Carlos used while analyzing the data and studying the top quark production at the colliders. Conversion of this parameter into the pole mass, which is the parameter required in these theoretical considerations and for the RGE, has uncertainties coming from hadronisation and fragmentation models, underlying event etc. These are typically non perturbative in character. Another way to extract the pole mass in a well-defined manner is to extract  $M_t^{\overline{MS}}$ , the mass of the top quark in the  $\overline{MS}$  scheme from the measurement of the top quark cross-sections at the Tevatron and the NNLO calculation of the same. The procedure to convert this mass to the pole mass  $M_t(M_t)$ , leads to uncertainties in  $M_t$  larger than the 0.7 GeV

taken in (18.45). This exercise, using the available information in 2012 led to an estimate of the pole mass for the top:

$$M_t^{pole} = 173.3 \pm 2.8 \text{ GeV} . \quad (18.46)$$

Compare this with the error of 0.7 GeV that was used in the estimate. The vacuum stability constraint now becomes  $M_h > 129.4 \pm 5.6 \text{ GeV}$  instead of the one in (18.45). This observation then can weaken the conclusion about the high scale upto which the SM remains valid without getting into conflict with stability. The future International Linear Collider (ILC) can measure the top mass  $M_t$  to a high accuracy of 100 MeV. What is more important is the fact that the determination of the  $t$  mass at the ILC comes directly from measurement of the  $t\bar{t}$  production cross-section in  $e^+e^-$  collisions, near the  $t\bar{t}$  threshold. This can be measured very accurately and has been computed theoretically to a high precision as well. This measurement can be converted into the pole mass in an unambiguous way. The figure below shows how such a precision measurement of the mass at the ILC can really shed light on whether the currently measured higgs mass points to the NEED of BSM physics at any *particular high scale*.



In the above figure, the bigger blue circle has been drawn assuming an LHC accuracy of  $t$  mass measurement of 1 GeV. However, a reduction of this error to about 500 MeV looks possible and is an active area of research. These kind of investigations are just the next logical step in our efforts to test the SM through a combination of the 'direct' and 'indirect' observations.

## 18.8 Discovery of the Higgs Boson

Because the Higgs boson decays very quickly, particle detectors cannot detect it directly. Instead the detectors register all the decay products (the decay signature) and from the data the decay process is reconstructed. If the observed decay products match a possible decay process (known as a decay channel) of a Higgs boson, this indicates that a Higgs boson may have been created. In practice, many processes may produce similar decay signatures. Fortunately, the SM precisely predicts the likelihood of each of these, and each known process, occurring. Because Higgs boson production in a particle collision is very rare (1 in 10 billion at the LHC), and many other possible collision events can have similar decay signatures, the data of hundreds of trillions of collisions needs to be analyzed and must "show the same picture" before a conclusion about the existence of the Higgs boson can be reached.

The first extensive search for the Higgs boson was conducted at the LEP at CERN in the 1990s. At the end of its service in 2000, LEP had found no conclusive evidence for the Higgs. This implied that if the Higgs boson were to exist it would have to be heavier than 114.4 GeV.

The search continued at Fermilab (US), where the Tevatron (the collider that discovered the top quark in 1995) had been upgraded for this purpose. There was no guarantee that the Tevatron would be able to find the Higgs, but it was the only supercollider that was operational since the LHC was still under construction and the planned Superconducting Super Collider had been cancelled in 1993 and never completed. The Tevatron was only able to exclude further ranges for the Higgs mass, and was shut down on 30 September 2011. The final analysis of the data excluded the possibility of a Higgs boson with a mass between 147 GeV and 180 GeV. In addition, there was a small (but not significant) excess of events possibly indicating a Higgs boson with a mass between 115 GeV and 140 GeV.

To find the Higgs boson, a more powerful accelerator with higher luminosity and advanced computing facilities were needed to process the vast amount of data (25 petabytes per year as of 2012) produced by the collisions. For the announcement of 4 July 2012, a new collider, LHC, was constructed at CERN with a planned eventual collision energy of 14 TeV (over seven times any previous collider) and over  $3 \times 10^{14}$  LHC proton-proton collisions were analyzed by the LHC Computing Grid, the world's largest computing grid (as of 2012), comprising over 170 computing facilities in a worldwide network

across 36 countries.

The LHC was designed specifically to be able to either confirm or exclude the existence of the Higgs boson. Built in a 27 km tunnel under the ground near Geneva originally inhabited by LEP, it was designed to collide two beams of protons, initially at energies of 3.5 TeV per beam (7 TeV total), or almost 3.6 times that of the Tevatron, and upgradeable to  $2 \times 7$  TeV (14 TeV total). As one of the most complicated scientific instruments ever built, its operational readiness was delayed for 14 months by a magnet quench event nine days after its inaugural tests, caused by a faulty electrical connection that damaged over 50 superconducting magnets and contaminated the vacuum system.

Data collection at the LHC finally commenced in March 2010. By December 2011 the two main particle detectors at the LHC, ATLAS and CMS, had narrowed down the mass range where the Higgs could exist to around 116 – 130 GeV (ATLAS) and 115 – 127 GeV (CMS). There had also already been a number of promising event excesses that had "evaporated" and proven to be nothing but random fluctuations. However, from around May 2011, both experiments had seen among their results, the slow emergence of a small yet consistent excess of  $\gamma$  and 4-lepton decay signatures and several other particle decays, all hinting at a new particle at a mass around 125 GeV. By around November 2011, the anomalous data at 125 GeV was becoming "too large to ignore" (although still far from conclusive), and the team leaders at both ATLAS and CMS each privately suspected they might have found the Higgs.

On November 28, 2011, at an internal meeting of the two team leaders and the director general of CERN, the latest analyses were discussed outside their teams for the first time, suggesting both ATLAS and CMS might be converging on a possible shared result at 125 GeV, and initial preparations commenced in case of a successful finding.

While this information was not known publicly at the time, the narrowing of the possible Higgs range to around 115 – 130 GeV and the repeated observation of small but consistent event excesses across multiple channels at both ATLAS and CMS in the 124 – 126 GeV region (described as "tantalizing hints" of around  $2 - 3\sigma$ ) were public knowledge with "a lot of interest".

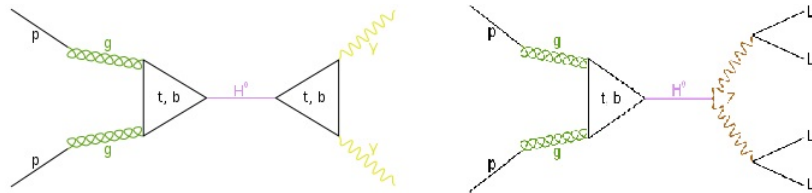
It was therefore widely anticipated around the end of 2011, that the LHC would provide sufficient data to either exclude or confirm the finding of a Higgs boson by the end of 2012, when their 2012 collision data (with slightly

higher 8 TeV collision energy) had been examined.

On 22 June 2012 CERN announced an upcoming seminar covering tentative findings for 2012, and shortly afterwards (from around 1 July 2012 according to an analysis of the spreading rumour in social media) rumours began to spread in the media that this would include a major announcement, but it was unclear whether this would be a stronger signal or a formal discovery. Speculation escalated to a "fevered" pitch when reports emerged that Peter Higgs, who proposed the particle, was to be attending the seminar, and that "five leading physicists" had been invited - generally believed to signify the five living 1964 authors - with Higgs, Englert, Guralnik, Hagen attending and Kibble confirming his invitation (died in 2011).

On 4 July 2012 both of the CERN experiments announced they had independently made the same discovery: CMS of a previously unknown boson with mass  $125.3 \pm 0.6$  GeV and ATLAS of a boson with mass  $126.0 \pm 0.6$  GeV. Using the combined analysis of two interaction types (known as 'channels'), both experiments independently reached a local significance of  $5\sigma$  - implying that the probability of getting at least as strong a result by chance alone is less than 1 in 3 million. When additional channels were taken into account, the CMS significance was reduced to  $4.9\sigma$ .

The figure below displays the Feynman diagrams showing the cleanest channels associated with the low-mass ( $\sim 125$  GeV) Higgs boson candidate observed by ATLAS and CMS at the LHC.



The dominant production mechanism at this mass involves two gluons from each proton fusing to a Top-quark Loop, which couples strongly to the Higgs field to produce a Higgs boson. *Left (Diphoton channel)*: Boson subsequently decays into two  $\gamma$ -ray photons by virtual interaction with a  $W$  boson loop or top quark loop. *Right (4-Lepton "golden channel")*: Boson emits two  $Z$  bosons, which each decay into two leptons (electrons, muons). Experimental analysis of these channels reached a significance of more than  $5\sigma$  in both



experiments.

On 31 July 2012, the ATLAS collaboration presented additional data analysis on the "observation of a new particle", including data from a third channel, which improved the significance to  $5.9\sigma$  (1 in 588 million chance of obtaining at least as strong evidence by random background effects alone) and mass  $126.0 \pm 0.4(\text{stat}) \pm 0.4(\text{sys})$  GeV, and CMS improved the significance to  $5\sigma$  and mass  $125.3 \pm 0.4(\text{stat}) \pm 0.5(\text{sys})$  GeV.

Following the 2012 discovery, it was still unconfirmed whether or not the 125 GeV particle was a Higgs boson. On one hand, observations remained consistent with the observed particle being the SM Higgs boson, and the particle decayed into at least some of the predicted channels. Moreover, the production rates and branching ratios for the observed channels broadly matched the predictions by the SM within the experimental uncertainties. However, the experimental uncertainties currently still left room for alternative explanations, meaning an announcement of the discovery of a Higgs boson would have been premature. To allow more opportunity for data collection, the LHC's proposed 2012 shutdown and 2013-14 upgrade were postponed by 7 weeks into 2013.

In January 2013, CERN director-general stated that based on data analysis to date, an answer could be possible 'towards' mid-2013, and the deputy chair of physics at Brookhaven National Laboratory stated in February 2013 that a "definitive" answer might require "another few years" after the collider's 2015 restart. In early March 2013, CERN Research Director stated that confirming spin-0 was the major remaining requirement to determine whether the particle is at least some kind of Higgs boson.

On 14 March 2013 CERN confirmed that: "CMS and ATLAS have compared a number of options for the spin-parity of this particle, and these all prefer no spin and even parity (two fundamental criteria of a Higgs boson consistent with the SM). This, coupled with the measured interactions of the new particle with other particles, strongly indicates that it is a Higgs boson". This also makes the particle the first elementary scalar particle to be discovered in nature.

In July 2017, CERN confirmed that all measurements still agree with the predictions of the SM, and called the discovered particle simply "the Higgs boson". As of April 2018, the LHC has continued to produce findings that confirm the 2013 understanding of the Higgs field and particle.

The LHC's experimental work since restarting in 2015 has included probing the Higgs field and boson in details, and confirming whether less common predictions were correct. In particular, exploration since 2015 has provided strong evidence of the predicted direct decay into fermions such as pairs of bottom quarks ( $3.6\sigma$ ), described as an *important milestone* in understanding its short lifetime and other rare decays, and also to confirm decay into pairs of tau leptons ( $5.9\sigma$ ). This was described by CERN as being "of paramount importance to establishing the coupling of the Higgs boson to leptons and represents an important step towards measuring its couplings to third generation fermions, the very heavy copies of the electrons and quarks, whose role in nature is a profound mystery". In July 2018, the ATLAS and CMS experiments reported observing the Higgs boson decay into a pair of bottom quarks, which makes up approximately 60% of all of its decays.

**Exercise 18.1:** Why the detection of the reaction  $\bar{\nu}_\mu + e^- = \bar{\nu}_\mu + e^-$  was the evidence of discovery of NC while  $\bar{\nu}_e + e^- = \bar{\nu}_e + e^-$  was not.

**Exercise 18.2:** In the OPAL experiment at LEP, the efficiencies for selecting  $W^+W^- \rightarrow l\nu q_1 q_2$  and  $W^+W^- \rightarrow q_1 \bar{q}_2 q_3 \bar{q}_4$  events were 83.8% and 85.9% respectively. After correcting for background, the observed numbers of  $l\nu q_1 q_2$  and  $q_1 \bar{q}_2 q_3 \bar{q}_4$  events were respectively 4192 and 4592. Determine the measured value of the  $W$ -boson hadronic branching ratio  $BR(W \rightarrow q\bar{q}')$  and its statistical uncertainty.

**Exercise 18.3:** A beam of negative muons can be stopped in matter because a muon may be: (a) transformed into an electron by emitting a photon; (b) absorbed by a proton, which goes into an excited state; (c) captured by an atom into a bound orbit about the nucleus.

**Exercise 18.4:** The straggling of heavy ions at low energy is mostly a consequence of: (a) finite momentum; (b) fluctuating state of ionization; (c) multiple scattering.

**Exercise 18.5:** The scattering of an energetic charged particle in matter is due mostly to interactions with: (a) electrons; (b) nuclei; (c) quarks.

**Exercise 18.6:** At low  $E/p$  the drift velocity of electrons in gases follows the relation  $v \propto E/p$ . This can be explained by the fact that: (a) the electrons each gains an energy  $eE \int ds$ ; (b) the electrons thermalize completely in inelastic encounters with the gas molecules; (c) the cross section is independent of electron velocity.

## Part VII

# Lecture – Introduction to the Standard Model



## Chapter 19

# Gauge Interactions

Now we are ready to start study *the Standard Model of Particle Physics* (SM), which (except for gravity) appears to be the theory which explains our universe. To state the SM in the simplest possible terms, it is:

A Yang-Mills (Gauge) Theory with Gauge Group  $SU(3) \times SU(2) \times U(1)_Y$  with 15 Weyl fermions, in three copies, of the representation  $(3, 2, 1/6) \oplus (1, 2, -1/2) \oplus (\bar{3}, 1, -2/3) \oplus (\bar{3}, 1, 1/3) \oplus (1, 1, 1)$  (corresponding to the left quark and lepton doublets, right up-type and down-type quarks and right charged leptons, respectively) and a single copy of a complex scalar field in the representation  $(1, 2, -1/2)$ .

### 19.1 Gauging the Symmetry

First let us mention *the Gauge Principle*, which is the basis of the theoretical description of three of the fundamental interactions, strong, weak and electromagnetic, among the quarks, leptons and the force carrying gauge bosons. QED is the first gauge theory to be established. We therefore can begin our discussion of gauge theories by looking at QED.

Consider the free Dirac Lagrangian, which is invariant under the global  $U(1)$  transformation

$$\psi \rightarrow e^{i\alpha}\psi . \tag{19.1}$$

We are going to make this symmetry *Local*, so that  $\alpha$  depends on spacetime,  $\alpha = \alpha(x^\mu)$ , and then try to force the Lagrangian to maintain its invariance under the local  $U(1)$  transformation. Making a global symmetry local is referred to as *Gauging the Symmetry*.

We start by making the local  $U(1)$  transformation:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \quad \rightarrow \quad \mathcal{L} = \bar{\psi}e^{-i\alpha(x)}(i\gamma^\mu\partial_\mu - m)e^{i\alpha(x)}\psi, \quad (19.2)$$

and because the differential operators will now act on  $\alpha(x)$  as well as  $\psi$ , we get extra terms:

$$\begin{aligned} \mathcal{L} &\rightarrow \bar{\psi}e^{-i\alpha(x)}(i\gamma^\mu\partial_\mu - m)e^{i\alpha(x)}\psi = \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \bar{\psi}\gamma^\mu\psi\partial_\mu\alpha(x) = \\ &= \bar{\psi}[i\gamma^\mu\partial_\mu - m - \gamma^\mu\partial_\mu\alpha(x)]\psi. \end{aligned} \quad (19.3)$$

If we want to demand that  $\mathcal{L}$  still be invariant under this local  $U(1)$  transformation, we must find a way of canceling the  $\bar{\psi}\gamma^\mu\psi\partial_\mu\alpha(x)$  term. We do this in the following way.

Define some arbitrary field  $A_\mu$  which under the  $U(1)$  transformation  $e^{i\alpha(x)}$  transforms according to

$$A_\mu \quad \rightarrow \quad A_\mu - \frac{1}{q}\partial_\mu\alpha(x). \quad (19.4)$$

We call  $A_\mu$  the *Gauge Field* for reasons that will be clear soon, and  $q$  is a constant we have included for later convenience. We introduce  $A_\mu$  by replacing the standard derivative  $\partial_\mu$  with the *Covariant Derivative*,

$$D_\mu \equiv \partial_\mu + iqA_\mu. \quad (19.5)$$

To say that a particle 'carries charge' mathematically means that it has the corresponding term in its covariant derivative. So, if a particle's covariant derivative is equal to the normal differential operator  $\partial^\mu$ , then the particle has no charge, and it will not interact with anything. But if it carries charge, it will have a term corresponding to that charge in its covariant derivative. This will become clearer as we proceed.

So, our Lagrangian is now

$$\begin{aligned} \mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi &= \bar{\psi}[i\gamma^\mu(\partial_\mu + iqA_\mu) - m]\psi = \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m - q\gamma^\mu A_\mu)\psi. \end{aligned} \quad (19.6)$$

Under the local  $U(1)$  we have

$$\begin{aligned}
\mathcal{L} &= \bar{\psi} e^{-i\alpha(x)} \left\{ i\gamma^\mu \partial_\mu - m - q\gamma^\mu \left[ A_\mu - \frac{1}{q} \partial_\mu \alpha(x) \right] \right\} e^{i\alpha(x)} \psi = \\
&= \bar{\psi} [i\gamma^\mu \partial_\mu - m - \gamma^\mu \partial_\mu \alpha(x) - q\gamma^\mu A_\mu + \gamma^\mu \partial_\mu \alpha(x)] \psi = \quad (19.7) \\
&= \bar{\psi} (i\gamma^\mu \partial_\mu - m - q\gamma^\mu A_\mu) \psi = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi .
\end{aligned}$$

So, the addition of the field  $A_\mu$  has indeed restored the  $U(1)$  symmetry. Notice that now it is not only invariant under this local  $U(1)$ , but also still under the global  $U(1)$  we started with, with the same conserved  $U(1)$  current

$$j^\mu = \bar{\psi} \gamma^\mu \psi . \quad (19.8)$$

This allows us to rewrite the Lagrangian as

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi - qj^\mu A_\mu . \quad (19.9)$$

But we have a problem. If we want to know what the dynamics of  $A_\mu$  will be, we naturally take the variation of the Lagrangian with respect to  $A_\mu$ . But because there are no derivatives of  $A_\mu$ , the Euler-Lagrange equation is

$$\frac{\partial \mathcal{L}}{\partial A_\mu} = -q\bar{\psi} \gamma^\mu \psi = 0 . \quad (19.10)$$

But

$$-q\bar{\psi} \gamma^\mu \psi = -qj^\mu . \quad (19.11)$$

So the equation of motion for  $A_\mu$  says that the current vanishes, or that

$$j^\mu = 0 , \quad (19.12)$$

and so the Lagrangian is reduced back to the free Dirac lagrangian, which was not invariant under the local  $U(1)$ .

We can state this problem in another way. All physical fields have some sort of dynamics. If they don't then they are a constant background field that never changes and does nothing. As it is written, equation (19.9) has a field  $A_\mu$ , but it has no kinetic term, and therefore no dynamics. So, to fix this problem we must include some sort of dynamics, or kinetic terms, for  $A_\mu$ .

For an arbitrary field  $A_\mu$ , the appropriate gauge-invariant kinetic term is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} , \quad (19.13)$$

where

$$F^{\mu\nu} \equiv \frac{i}{q}[D^\mu, D^\nu] \quad (19.14)$$

and  $q$  is the constant of proportionality introduced in the transformation of  $A_\mu$  in equation (19.4).  $D^\mu$  is the covariant derivative defined in (19.5).

Writing out (19.14) and using an arbitrary test function  $f(x)$ ,

$$\begin{aligned} F^{\mu\nu} f(x) &= \frac{i}{q}[D^\mu, D^\nu]f(x) = \\ &= \frac{i}{q}[(\partial^\mu + iqA^\mu)(\partial^\nu + iqA^\nu) - (\partial^\nu + iqA^\nu)(\partial^\mu + iqA^\mu)]f = \\ &= \frac{i}{q}[\partial^\mu\partial^\nu f + iq\partial^\mu(A^\nu f) + iqA^\mu\partial^\nu f - q^2 A^\mu A^\nu f - \\ &\quad - \partial^\nu\partial^\mu f + iq\partial^\nu(A^\mu f) + iqA^\nu\partial^\mu f - q^2 A^\nu A^\mu f] = \\ &= (\partial^\mu A^\nu - \partial^\nu A^\mu + iq[A^\mu, A^\nu])f(x). \end{aligned} \quad (19.15)$$

But for each value of  $\mu$ ,  $A^\mu$  is a scalar function, so the commutator term vanishes, leaving (dropping the test function  $f$ )

$$F^{\mu\nu} = \frac{i}{q}[D^\mu, D^\nu] = \partial^\mu A^\nu - \partial^\nu A^\mu. \quad (19.16)$$

So, writing out the entire Lagrangian we have

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (19.17)$$

Finally, because  $A^\mu$  is obviously a physical field, we can naturally assume that there is some source term causing it, which we simply call  $J^\mu$ . This makes our final Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu. \quad (19.18)$$

So we started with a Lagrangian for a spin-1/2 particle, which had a global  $U(1)$  symmetry. Then, we promote the  $U(1)$  symmetry to a local symmetry (we gauged the symmetry), and then imposed what we had to impose to get a consistent theory. The gauge field  $A_\mu$  was forced upon us, and the form of the kinetic term for  $A_\mu$  is demanded automatically by geometric considerations.



The gauge symmetry in electromagnetism is a sort of remnant of the much deeper and more fundamental  $U(1)$  structure of the theory. In other words, we started with a non-interacting particle, and by specifying  $U(1)$  we have created a theory with not only that same particle, but also electromagnetism. The  $A_\mu$  field, which upon quantization will be the photon, is a direct consequence of the  $U(1)$ .

Theories of this type, where we generate forces by specifying a Lie group, are called *Gauge Theories*, or *Yang-Mills Theories*.

## 19.2 Non-Abelian Case

We are now ready to generalize what we did in section 19.1 to an arbitrary Lie group. Consider a Lagrangian  $\mathcal{L}$  with  $N$  scalar (or spinor) fields  $\phi_i$  ( $i = 1, \dots, N$ ) that is invariant under a continuous  $SO(N)$  or  $SU(N)$  symmetry,

$$\phi_i \rightarrow U_{ij}\phi_j, \quad (19.19)$$

where  $U_{ij}$  is an  $N \times N$  matrix of  $SO(N)$  or  $SU(N)$ .

In section 19.1, we saw that if the group is  $U(1)$ , gauging it demands the introduction of the gauge field  $A^\mu$  to preserve the symmetry, which shows up in the covariant derivative  $D_\mu = \partial_\mu - ieA_\mu$ . To say a field carries some sort of charge means that it has the corresponding term in its covariant derivative. We then added a kinetic term for  $A^\mu$  as well as an external source  $J^\mu$ . Then, higher order interaction terms can be included in whatever way is appropriate for the theory.

To generalize this, let's say for the sake of concreteness that our Lie group is  $SU(N)$ . An arbitrary element of  $SU(N)$  is  $e^{ig\theta^a(x)T^a}$ , where  $g$  is a constant we have added for later convenience,  $\theta^a$  are the  $N^2 - 1$  parameters of the group (see the Exercise 14.3), and the  $T^a$  are the generator matrices for the group. Notice that we have gauged the symmetry (in that  $\theta(x)$  is a function of spacetime).

By definition, we know that the generators  $T^a$  will obey the commutation relations

$$[T^a, T^b] = if_{abc}T^c, \quad (19.20)$$

where  $f_{abc}$  are the *structure constants* of the group.

When gauging the  $U(1)$  in section 19.1, the transformation of the gauge field was given by equation (19.4). For the more general transformation (19.19), the gauge field transforms according to

$$A^\mu \rightarrow U(x)A^\mu U^\dagger(x) + \frac{i}{g}U(x)\partial^\mu U^\dagger(x), \quad (19.21)$$

where we have removed the initial notation and it is understood that matrix multiplication is being discussed. If  $U(x)$  is an element of  $U(1)$  (so it is  $e^{ig\theta(x)}$ ), then this transformation reduces to

$$\begin{aligned} A^\mu &\rightarrow e^{ig\theta(x)}A^\mu e^{-ig\theta(x)} + \frac{i}{g}e^{ig\theta(x)}[-ig\partial^\mu\theta(x)]e^{-ig\theta(x)} = \\ &= A^\mu + \partial^\mu\theta(x). \end{aligned} \quad (19.22)$$

For general  $SU(N)$ , the  $U$ 's are elements of a non-Abelian group, and the  $A_\mu$ 's are matrices of the same size.

Generalizing, we find that a general element of the  $SU(N)$  is (changing notation slightly)

$$U(x)e^{-ig\Gamma^a(x)T^a}, \quad (19.23)$$

with  $N^2 - 1$  real parameters  $\Gamma^a$ . We then build the covariant derivative in the exact same way as in equation (19.5) by adding a term proportional to the gauge field

$$D_\mu = \mathbb{I}^{N \times N} \partial_\mu - igA_\mu \quad (19.24)$$

(Remember that each component of  $A_\mu$  is an  $N \times N$  matrix. They were scalars for  $U(1)$  because  $U(1)$  is a  $1 \times 1$  matrix). Or, acting on the fields, the covariant derivative is

$$(D_\mu\phi)_j = \partial_\mu\phi_j(x) - ig[A_\mu(x)]_{jk}\phi_k(x), \quad (19.25)$$

where  $k$  is understood to be summed on the last term. It will be understood from now on that the normal partial derivative term (the first term) has an  $N \times N$  identity matrix multiplied by it.

Then, just as in (19.14), we have the field strength

$$F_{\mu\nu}(x) \equiv \frac{i}{g}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu], \quad (19.26)$$

where the commutator term doesn't vanish for arbitrary Lie group as it did for Abelian  $U(1)$ .

Recall that for  $U(1)$ ,  $F_{\mu\nu}$  is invariant under the gauge transformation on its own, because the commutator term vanishes. In general, however, the commutator term does not vanish, and we must therefore be careful in writing down the correct kinetic term. It turns out that the correct choice is

$$\mathcal{L}_{\text{kinetic}} = -\frac{1}{2}\text{Tr}(F_{\mu\nu}F^{\mu\nu}) . \quad (19.27)$$

So, starting with a non-interacting Lagrangian that is invariant under the global  $SU(N)$ , we can gauge the  $SU(N)$  to create a theory with a gauge field (or synonymously a 'force carrying' field)  $A^\mu$ , which is an  $N \times N$  matrix. So, every Lie group gives rise to a particular gauge field (which is a force carrying particle, like the photon), and therefore a particular force.

For this reason, we discuss forces in terms of Lie groups, or synonymously *Gauge Groups*. Each group defines a force, e.g.  $U(1)$  represents the electromagnetic force (as we have seen in section 19.1), while  $SU(2)$  describes the weak force, and  $SU(3)$  describes the strong color force.

## 19.3 Representations of Gauge Groups

As we discussed above, given a set of structure constants  $f_{abc}$ , which define the Lie algebra of some Lie group, we can form a representation of that group, which we denote  $R$ . So,  $R$  will be a set of  $D(R) \times D(R)$  matrices, where  $D$  is the dimension of the representation  $R$ . We then call the generators of the group (in the representation  $R$ )  $T_R^a$ , and they naturally obey

$$[T_R^a, T_R^b] = if_{abc}T_R^c . \quad (19.28)$$

One representation which exists for any of the groups we have considered is the representation of  $SO(N)$  or  $SU(N)$  consisting of  $N \times N$  matrices. We denote this as the *Fundamental Representation*. Clearly, the fundamental representations of  $SO(2)$ ,  $SO(3)$ ,  $SU(2)$ , and  $SU(3)$  are the  $2 \times 2$ ,  $3 \times 3$ ,  $2 \times 2$ , and  $3 \times 3$  matrix representations, respectively. So, the fundamental representation of  $SU(2)$  will be denoted 2, and the generators for  $SU(2)$  in the fundamental representation will be denoted  $T_2^a$ . Obviously, the fundamental representation of  $SU(3)$  will be 3 with generators  $T_3^a$ .

Furthermore, let's say we have some arbitrary representation generated by  $T_R^a$ , obeying (19.28). We can take the complex conjugate of the commutation

relations to get

$$[T_R^{*a}, T_R^{*b}] = -if_{abc}T_R^{*c} . \quad (19.29)$$

So, notice that if we define the new set of generators

$$T_R'^a \equiv -T_R^{*a} , \quad (19.30)$$

then the  $T_R'^a$  will obey the correct commutation relations, and will therefore form a representation of the group as well. If it turns out that

$$T_R'^a = -(T_R^a)^* = T_R^a , \quad (19.31)$$

or if there is some unitary similarity transformation

$$T_R^a \rightarrow U^{-1}T_R^aU \quad (19.32)$$

such that

$$T_R'^a = -(T_R^a)^* = T_R^a , \quad (19.33)$$

then we call the representation *Real*, and the complex conjugate of the  $T_R^a$ 's is the same representation. However, if no such transformation exists, then we have a new representation, called the *Complex Conjugate* representation to  $R$ , or the *Anti- $R$*  representation, which we denote  $\bar{R}$ . For example, there is the fundamental representation of  $SU(3)$ , denoted  $\mathbf{3}$ , generated by  $T_3^a$ , and then there is the anti-fundamental representation  $\bar{\mathbf{3}}$ , generated by  $T_3^a$ .

The representations of a group which will be important to us are the fundamental, anti-fundamental, and adjoint.

**Exercise 19.1:** Construct the covariant derivative for scalar fields in the adjoint representation.

**Exercise 19.2:** Consider two sets of scalar fields,  $\phi_1$  and  $\phi_2$ , which transform as vector representations under the  $O(n)$  group and construct the covariant derivative for  $\phi_1$ .

**Exercise 19.3:** Show that any Hermitian  $3 \times 3$  matrix can be written as a linear combination of the unit matrix and the eight Gell-Mann matrices.

**Exercise 19.4:** Suppose it is given an eight dimensional vector:  $G_a^\nu$  ( $a = 1, 2, \dots, 8$ ). Write out the form of the covariant derivatives using the matrix notation.

## Chapter 20

# The Unification Scheme

### 20.1 Fermi's Theory for Weak Interactions

Fermi's theory of  $\beta$  decay was the blueprint of the early theoretical description of the weak interactions, which are responsible not just for the radioactive  $\beta$  decays of nuclei, but also for the strangeness conserving and strangeness changing weak decays of the mesons and baryons. This culminated in the famous  $V - A$  theory of weak interactions, according to which the muon decay,

$$\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e, \quad (20.1)$$

for example, could be described by an effective Hamiltonian

$$\mathcal{H}_{eff}^{\mu \text{ decay}} = -\frac{G_\mu}{\sqrt{2}} \left[ J_{\nu e}^{\rho+\dagger} J_{\nu\mu,\rho}^+ + h.c. \right], \quad (20.2)$$

where *the Charged Current* (CC) is defined as:

$$J_\rho^+ = \bar{\psi}_1 \gamma^\rho (1 - \gamma_5) \psi_2 \equiv J_\rho^{CC}. \quad (20.3)$$

In the same way, the  $\beta$  decay of the neutron could be described by an effective interaction given by

$$\mathcal{H}_{eff}^{\beta \text{ decay}} = -\frac{G_F}{\sqrt{2}} \left[ J_{e\nu}^{\rho+\dagger} J_{pn,\rho}^+ + h.c. \right], \quad (20.4)$$

with

$$J_{pn}^{\rho+} = \bar{\psi}_p (1 - 1.26\gamma_5) \gamma^\mu \psi_n. \quad (20.5)$$

Generic four fermion interaction responsible for the weak processes (left panel) and the basic process describing the  $\beta$  decay (right panel).



It was established that when written in terms of the quarks which make up the mesons and baryons, all the weak processes could be described in terms of a four fermion, current-current interaction depicted in left panel of the Figure above, which shows a transition  $1 \rightarrow \bar{2} + 3 + 4$ . For example, the basic transition describing the  $n$  decay

$$n(udd) \rightarrow p(uud) + e^- + \bar{\nu}_e, \quad (20.6)$$

is given by the current-current interaction depicted in the right panel.

The crux of  $V - A$  theory is that only the left chiral fermions are involved in this weak Hamiltonian. The effective Hamiltonian is then written as:

$$\begin{aligned} \mathcal{H}_{eff}^{4fermion} &= -\frac{G_\mu}{\sqrt{2}} [J_{24}^{\mu\dagger} J_{31,\mu}^+ + h.c.] = \\ &= -4\frac{G_\mu}{\sqrt{2}} [(\bar{\psi}_{3L}\gamma^\mu\psi_{1L})(\bar{\psi}_{4L}\gamma_\mu\psi_{2L}) + h.c.] . \end{aligned} \quad (20.7)$$

The appearance

$$\psi_L = \frac{1}{2}(1 - \gamma_5)\psi \quad (20.8)$$

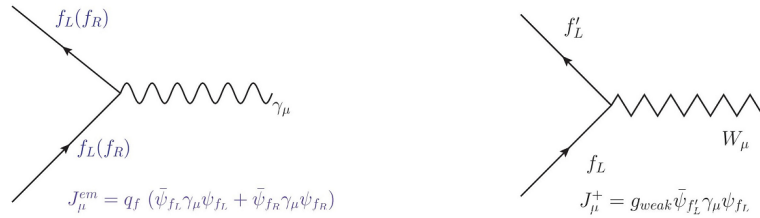
in (20.7), indicates that only left chiral fermions are involved in this weak CC. As we will see later, it is this fact that decides the representation of the  $SU(2)_L$  gauge group to which the various fermion fields belong.

## 20.2 Decay Constants

We understand the electromagnetic interaction in terms of the electromagnetic current

$$J_\mu^{em} = \bar{\psi}_L\gamma_\mu\psi_L + \bar{\psi}_R\gamma_\mu\psi_R \quad (20.9)$$

and the electromagnetic field  $A_\mu$ . The corresponding vertex is depicted in the left panel of the Figure below. The form of effective Hamiltonian (20.7) means that one can similarly think of the weak current  $J_\mu^\pm$  coupled to a charged gauge boson  $W_\mu^\pm$ . The basic transition brought about by the CC could then be depicted as shown in the right panel of the Figure.



The electromagnetic charge of  $f'$  differs from that of  $f$  by one unit and in case  $f$  is strange quark, the strangeness changes by one unit as well. In that case this current indicates a transition which brings about  $\Delta S = \Delta Q = 1$ , where  $S$  and  $Q$  stand for the strangeness and the electromagnetic charge respectively. While the decay of a neutron  $n$  involves the current  $J_{ud}^{\mu+}$ , the decay of  $\Lambda$  for example, involves the current  $J_{us}^{\mu+}$ . The strength of the four-fermion interaction is then decided by  $g_{weak}$  of the Figure above.

Experimentally measured values of  $G_\mu$  and  $G_F$ , of (20.2) and (20.4), were somewhat different from each other, though very close,  $G_F \sim 0.98 G_\mu$ . For the effective Hamiltonian for  $\Lambda$  decay the corresponding coefficient  $G_\Lambda$  was yet again different from both  $G_\mu$  and  $G_F$ , being  $0.20 G_\mu$ . The near equality between  $G_\mu$  and  $G_F$  was an indication that the vector current was not affected by the strong interactions of  $n$  and  $p$  and is the same for  $e^- \rightarrow \nu$  transition as for  $n \rightarrow p$ . This was called *the conserved vector current hypothesis*.

## 20.3 Quark Mixing

It was Cabibbo's observation that conservation hypothesis could be consistent with a completely universal charged weak current, i.e. a current which has the same strength for the leptons as well as the quarks and also for  $\Delta S = 0$  and  $\Delta S = 1$  alike, if in case of quarks, the basic CC in the Figure

of the previous section describes a transition

$$f' = u, \quad f = d' = d \cos \theta_c + s \sin \theta_c, \quad (20.10)$$

with  $\sin \theta_c \sim 12^\circ$ . This means that the interaction eigenstate  $d'$  is a linear combination of the mass eigenstates  $d$  and  $u$ . Clearly, the orthogonal combination

$$s' = -d \sin \theta_c + s \cos \theta_c, \quad (20.11)$$

is an interaction eigenstate coupling with a  $W^\pm$  and a new quark with charge  $+2/3$ . This thus indicates existence of the fourth quark: the charm quark  $c$ .

As we will see later existence of the fourth quark ensures flavor conservation of the weak *Neutral Currents* (NC) at tree level automatically. This then helps one understand the experimentally observed suppression of the flavor changing NCs which will be discussed in detail later. Thus the states to be identified with the interaction eigenstates would be:

$$\begin{pmatrix} u' \\ d' \end{pmatrix} = \begin{pmatrix} u \\ d \cos \theta_c + s \sin \theta_c \end{pmatrix}; \quad \begin{pmatrix} c' \\ s' \end{pmatrix} = \begin{pmatrix} c \\ -d \sin \theta_c + s \cos \theta_c \end{pmatrix}. \quad (20.12)$$

At this point let us also mention one more feature of the phenomenology of quark mixing which will be relevant later. In fact, the physics of the  $K_0$ ,  $\bar{K}_0$  mesons not only revealed the existence of suppressed nature of the flavor changing NCs but also CP violation in  $K_0 - \bar{K}_0$  system. This CP violation can also be understood as coming from the above quark-mixing but only if the mixing matrix involves a phase. For this to be possible we have to have at least three generations of quarks. This was noted by Kobayashi-Maskawa. This makes it possible to understand the CP violation observed in the neutral meson system, in the context of a gauge theory of EW interactions, in terms of the mixing in the quark sector. However, this requires existence of at least three generations. Thus, one sees that in some sense, the need to understand the observed phenomenology of flavor changing NCs and CP violation, in the framework of a gauge theory, the existence of the  $c$  and the  $t$  quark was predicted.

For future reference note that the connection between the mass eigenstates  $u, d, c, s, t$  and  $b$  and the interaction eigenstates  $u', d', c', s', t'$  and  $b'$  is



given by  $u' = u$ ,  $c' = c$ ,  $t' = t$  and

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (20.13)$$

where  $V_{ud}$  etc. in (20.13) are elements of so called Kobayashi-Maskawa matrix. This describes the interaction eigenstates in terms of the mass eigenstates.

## 20.4 Intermediate Bosons

Let us note that the same four fermion interaction that describes the decay (20.1) can also describe, for example, the scattering processes such as

$$\nu_\mu + e^- \rightarrow \nu_e + \mu^-, \quad (20.14)$$

corresponding to  $1 = e^-$ ,  $2 = \nu_\mu$ ,  $3 = \nu_e$  and  $4 = \mu^-$  in the left panel of the Figure in Sec. 20.1. The same effective Hamiltonian as in (20.7) then also describes this scattering process as well. If one calculates the total cross-section one gets,

$$\sigma_{\text{tot}} = \frac{G_\mu^2 s}{\pi} = \frac{2G_\mu^2 m_e E_{\nu_\mu}}{\pi}. \quad (20.15)$$

This linear rise of scattering cross-section with  $s$ , the square of the center of mass energy or alternatively  $E_{\nu_\mu}$ , is a reflection of the 'point-like' nature of the Fermi interaction of (20.7).

It can be seen, by doing a partial wave analysis of the scattering amplitude, that this behavior implies violation of unitarity when

$$\sqrt{s} \geq 300 \text{ GeV}. \quad (20.16)$$

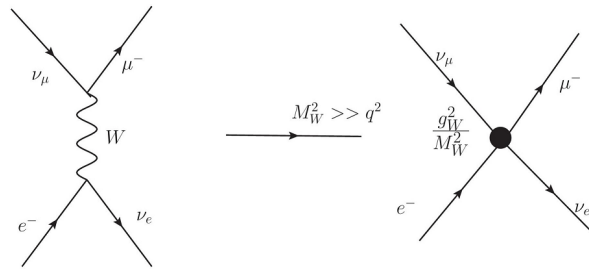
Of course, in practical terms it corresponds to a

$$E_{\nu_\mu} \geq 10^8 \text{ GeV} \quad (20.17)$$

and hence not very relevant. However, it is the principle that matters.

A cure to the unitarity problem of the current-current interaction was offered by Schwinger, who postulated the existence of a massive, charged boson

(called the weak-boson  $W^\pm$ ). This is the same  $W^\pm$  we have already introduced while writing the weak vertex in the Figure of Sec. 20.2. Thus the point interaction of (20.7) can be understood as an exchange of a  $W^\pm$  boson, in the limit of the said mass  $M_W$  being much bigger than all the energies in the system. This is depicted in the Figure below.



The observed short range of the weak force causing the  $\beta$  decay, indicated that the  $W^\pm$  boson is massive, unlike the photon mediating the electromagnetic interaction which is massless. To summarize, we see that the requirement that unitarity bound be respected, indicates the existence of a massive charged vector boson  $W^\pm$  and the four-fermion weak interactions can be understood as caused by an exchange of this massive boson. The 'massive' nature of the exchanged boson was also consistent with the observed 'short' range of the weak interactions. However, if it is a gauge boson, then the massive nature will also break gauge invariance! Further, the massive nature of the gauge boson causes problems such as bad high energy behavior of scattering amplitudes as well as non renormalisability of the theory. How a massive gauge boson is to be accommodated in the framework of a gauge theory is going to be the topic of discussion in the next lecture.

**Exercise 20.1:** A muon at rest lives  $10^{-6}$  sec and its mass is 100 MeV. How energetic must a muon be to reach the earth's surface if it is produced high in the atmosphere (say  $\sim 10^4$  m up)?

**Exercise 20.2:** Suppose to a zeroth approximation that the earth has a 1-gauss magnetic field pointing in the direction of its axis, extending out to  $10^4$  m. How much, and in what direction, is a muon of energy  $E$  normally incident at the equator deflected by the field?

**Exercise 20.3:** Estimate the attenuation (absorption/scattering) of a beam of 50-keV X-rays in passage through a layer of human tissue (no bones!) one

centimeter thick.

**Exercise 20.4:** Suppose there exists a neutral particle  $\psi$ , which is the strong isospin singlet. What value you expect for the coupling constants of this particle with protons and neutrons?



## Chapter 21

# Particle Content and Currents

In this chapter we want to write down the particle content for the EW (Glashow-Salam-Weinberg) model along the all interactions among them.

As we mentioned above the gauge group for the EW model is  $SU(2)_L \times U(1)_Y$ . The subscript  $L$  means that the gauge transformations corresponding to this gauge group are non-trivial only for the left (chiral) fermions and the right fermions remain unchanged under it. The direct product means that these two groups are independent, i.e. the left fermions belonging to a given representation of  $SU(2)_L$  will all have the same value of the charge under  $U(1)_Y$ . Thus only the left chiral fermions belong to the nontrivial representation of the  $SU(2)_L$  group and the right chiral fermions are singlets under the  $SU(2)_L$  gauge group. Therefore these have no interactions with the gauge bosons corresponding to the  $SU(2)_L$  gauge group.

### 21.1 Gauge Bosons

For the  $SU(2)$  group, each representation is labelled by two quantum numbers  $T_L$  and  $T_{3L}$ , where  $T_L$  takes integral or half integer values (0, 1/2, 1, 3/2... etc.) and for a given  $T_L$ ,  $T_{3L}$  takes values from  $-T_L$  to  $+T_L$  in steps of 1. Thus number of fields belonging to representation labelled by  $T_L$  is then  $2T_L + 1$ . For singlet representation  $T_L = 0$  and for the doublet it is 1/2. Thus a doublet of  $SU(2)_L$  contains two members with  $T_{3L} = \pm 1/2$ . The gauge bosons belong to the  $T = 1$  representation (called the adjoint

representation) and hence they are three in number called  $W_\mu^a$  ( $a = 1, 2, 3$ ).

The  $U(1)_Y$  gauge group has only one generator, like the QED. We denote the corresponding single gauge boson as  $B_\mu$ . The corresponding current is  $J_\mu^Y$  and the charge is called *hypercharge*. The electromagnetic charge of a fermion is independent of its chirality. On the other hand, the two left fermions of different electromagnetic charges have to have the same  $U(1)_Y$  charge. Thus it is clear that the  $U(1)_Y$  cannot be identified with  $U(1)_{em}$ , i.e. the hypercharge is different from the electromagnetic charge. Thus  $U(1)_{em}$  arises out of a linear combination of  $U(1)_Y$  and a  $U(1)$  subgroup of  $SU(2)_L$ .

The gauge groups, the corresponding spin-1 gauge bosons  $W_\mu^a$  and  $B_\mu$  and the couplings are indicated in the table below.

Gauge Group	Gauge Boson Fields	Coupling
$SU(2)_L$	$W_\mu^a$ ( $a = 1, 2, 3$ )	$g_2$
$U(1)_Y$	$B_\mu$	$g_1$

## 21.2 Fermions

If the left fermions belong to the doublet representation of  $SU(2)_L$ , the corresponding gauge CC,  $J_\mu^W$ , has the same form as the  $J_\mu^{CC}$  of (20.3), of the  $V - A$  current Lagrangian describing the charge changing weak interactions.

Let  $Y/2$  denote the charge of the fermion under the  $U(1)_Y$  gauge group. The corresponding transformation is given by

$$\psi \rightarrow e^{-i(g_1 Y/2)\alpha_Y(x)} \psi, \quad (21.1)$$

where as for a  $SU(2)_L$  doublet the gauge transformation is given by

$$\Psi = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \rightarrow \Psi' = e^{-ig_2(\tau^a/2)\alpha^a(x)} \Psi. \quad (21.2)$$

$f_1$  and  $f_2$  are the  $T_{3L} = \pm 1/2$  members of this doublet  $\Psi$  respectively.  $\tau^a/2$  are the generators  $T^a$  for the 2-dimensional fundamental representation.

The fermion content of the EW model is written in the table below.

Quarks	Leptons
$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$ $u_R, c_R, t_R$ $d_R, s_R, b_R$ + anti-quarks	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$ $e_R, \mu_R, \tau_R$ + anti-leptons

All the left chiral fermions belong to the doublet representation, with the up-type quarks and neutrinos having  $T_{3L} = 1/2$  and down-type quarks and negatively charged leptons having  $T_{3L} = -1/2$ . Note that according to this table there are no right handed neutrinos in the particle spectrum of the SM. The colour gauge group  $SU(3)_c$  commutes with the EW gauge group:  $SU(2)_L \times U(1)_Y$ . Hence the EW interactions of a quark are independent of its colour. Therefore we suppress here the colour index.

### 21.3 Neutral Bosons Mixing

As already discussed  $U(1)_{em}$  is a linear combination of  $U(1)_Y$  and a  $U(1)$  subgroup of  $SU(2)$ . This is really the essence of EW unification and is embodied in Glashow's observation:

$$Q_f = T_{3L} + \frac{Y}{2}. \quad (21.3)$$

Here  $Q_f$  is the electromagnetic charge in units of  $|e|$ , where  $e$  is electron charge,  $T_{3L}$  and  $Y/2$  denote the  $SU(2)_L$  and  $U(1)_Y$  charges respectively. Writing the electromagnetic charge as a linear combination of  $T_{3L}$  and the hyper-charge  $Y$ , embodies the fact that the carrier of electromagnetic interactions, the photon  $A_\mu$  will appear as a linear combination of the neutral vector boson  $W_\mu^3$  and the  $U(1)_Y$  gauge boson  $B_\mu$ . We can discuss this mixing without making any explicit reference to the Higgs sector. This is what we will do first and then summarize the details of the SSB. Note that the three gauge boson fields  $W_\mu^1, W_\mu^2, W_\mu^3$ : all couple only to left handed fermions and  $B_\mu$  couples to both the left handed and right handed fermions.  $B_\mu$  and  $W_\mu^3$  mix, giving one zero mass eigenstate  $\gamma$ . One then identifies the other one with a new neutral vector boson called  $Z$ .

Note here that one can discuss this simply at the level of currents which give interactions among matter and gauge bosons in terms of the gauge principle,

without making any reference to a specific model which will generate these mixing and masses. The essence of this mixing is to define two fields  $A_\mu$  and  $Z_\mu$  as a linear combination of  $B_\mu$  and  $W_\mu^3$  as:

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3, \quad Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^3. \quad (21.4)$$

Here,  $\theta_W$ , called *the weak mixing angle*, is just an arbitrary parameter denoting the mixing between the  $W_\mu^3$  and  $B_\mu$ .

## 21.4 Currents

To see how the electric charge  $e$  is related to  $g_1$ ,  $g_2$  and  $\sin \theta_W$ , let us construct the currents,  $J_\mu^W$  and  $J_\mu^Y$ , the way electromagnetic current was constructed. To do this we need to know the  $Y$  values for the different fermion fields written in the table above. Let us consider a single generation of leptons:  $e^-$ ,  $\nu_e$ . From (21.3) we see that the lepton doublet

$$\mathcal{L}_L^1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad (21.5)$$

has  $Y = -1$  and  $e_{1R} = e_R$  which is an  $SU(2)_L$  singlet has to have  $Y = -2$ . Let us indicate the three lepton doublets written in the last three columns of the fermions table by  $\mathcal{L}^i$  with  $i = 1, 2, 3$  respectively. We also use  $\mathcal{Q}_L^i$  with  $i = 1, 2, 3$  to indicate the quark doublets, as written in the first three columns of the same table. For the quark doublets  $\mathcal{Q}^i$  the hypercharge  $Y$  has value  $1/3$ . For all the right handed quarks the hypercharge is twice the quark charge and  $Y = 2Q_q$ , since the value of  $T_{3L}$  is zero for all the right handed fields.

To construct the physical currents of the EW model, let us start from the kinetic part of the Lagrangian for all fermions. For the  $SU(2)_L \times U(1)_Y$  gauge theory the partial derivative is to be replaced by the covariant derivative, which can be written in terms of the hyper charges for the fermions. For a fermion  $f$  which is a member of the doublet  $\Psi$  this is given by:

$$\partial_\mu \Psi_L \rightarrow D_\mu \Psi_L = \partial_\mu \Psi_L - i \frac{g_1 Y_\Psi}{2} B_\mu \Psi_L - i g_2 W_\mu^a \frac{\tau^a}{2} \Psi_L. \quad (21.6)$$

where  $\Psi_L = \mathcal{L}_L^i$ ,  $\mathcal{Q}_L^i$  and  $Y_\Psi$  is the hypercharge for the doublet  $\Psi$ . For the case of  $SU(2)_L$  singlets the covariant derivative is given by

$$D_\mu f_R = \partial_\mu f_R - i \frac{g_1 Y_{f_R}}{2} B_\mu f_R. \quad (21.7)$$



The kinetic terms for all the fermions can be written as:

$$\mathcal{L}_{\text{kin}} = \sum_{i=1}^3 [i\mathcal{L}_L^i \partial \mathcal{L}_L^i + ie_R^i \partial e_R^i + i\mathcal{Q}_L^i \partial \mathcal{Q}_L^i + iu_R^i \partial u_R^i + id_R^i \partial d_R^i] . \quad (21.8)$$

Since there are no right handed neutrinos in the strictest version of the SM, for the lepton sector the mass basis and interaction basis are the same. Using the expressions for the covariant derivative  $D_\mu$  of (21.6) and (21.7), along with (20.13), we find the interaction Lagrangian to be

$$\Delta\mathcal{L}_{\text{int}} = \frac{1}{2}g_1 J^{\mu Y} B_\mu + g_2 \left[ \frac{1}{2\sqrt{2}}(J^{\mu+}W_\mu^+ + J^{\mu-}W_\mu^-) + J^{\mu 3}W_\mu^3 \right] , \quad (21.9)$$

where:

$$\begin{aligned} J^{\mu+} &= 2 \left( \bar{\nu}_L^i \gamma^\mu e_L^i + \bar{u}_L^i \gamma^\mu \mathbf{V}_{ij} d_L^j \right) = (J^{\mu-})^\dagger , \\ J^{\mu Y} &= -\bar{\nu}_L^i \gamma^\mu \nu_L^i - \bar{e}_L^i \gamma^\mu e_L^i - 2\bar{e}_R^i \gamma^\mu e_R^i + \frac{1}{3}\bar{u}_L^i \gamma^\mu u_L^i + \\ &\quad + \frac{1}{3}\bar{d}_L^i \gamma^\mu d_L^i + \frac{4}{3}\bar{u}_R^i \gamma^\mu u_R^i - \frac{2}{3}\bar{d}_R^i \gamma^\mu d_R^i , \\ J^{\mu 3} &= \frac{1}{2}\bar{\nu}_L^i \gamma^\mu \nu_L^i - \frac{1}{2}\bar{e}_L^i \gamma^\mu e_L^i + \frac{1}{2}\bar{u}_L^i \gamma^\mu u_L^i - \frac{1}{2}\bar{d}_L^i \gamma^\mu d_L^i , \\ W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) , \end{aligned} \quad (21.10)$$

where  $\mathbf{V}_{ij}$  denotes the Kobayashi-Maskawa matrix.

The couplings in (21.9) must be rewritten so that one linear combination of  $B_\mu$ ,  $W_\mu^3$  couples to the electromagnetic current and an orthogonal one couples to  $J^{\mu 3}$ . For this purpose we may ignore the terms in  $\Delta\mathcal{L}$  depending on  $W^\pm$ . For the remaining part, we may think of the physical fields  $A_\mu$ ,  $Z_\mu$  as the result of a rotation in the  $B_\mu$ ,  $W_\mu^3$  plane, as already discussed in (21.4). We write the inverse rotation:

$$W_\mu^3 = \cos\theta_W Z_\mu + \sin\theta_W A_\mu , \quad B_\mu = -\sin\theta_W Z_\mu + \cos\theta_W A_\mu . \quad (21.11)$$

Inserting into the Lagrangian (21.9), we find:

$$\begin{aligned} \Delta\mathcal{L}(B_\mu, W_\mu^3) &= \left[ \frac{1}{2}g_1 \cos\theta_W J^{\mu Y} + g_2 \sin\theta_W J^{\mu 3} \right] A_\mu + \\ &\quad + \left[ -\frac{1}{2}g_1 \sin\theta_W J^{\mu Y} + g_2 \cos\theta_W J^{\mu 3} \right] Z_\mu . \end{aligned} \quad (21.12)$$

The expression in the first bracket in (21.12) must be equal to  $eJ^{\mu\text{em}}A_\mu$  where  $e$  is the unit of electric charge and  $J^{\mu\text{em}}$  is given by an expression for all the charged fermions and can be written as

$$\begin{aligned} J_\mu^{\text{em}} &= -\bar{e}_L^i \gamma_\mu e_L^i - \bar{e}_R^i \gamma_\mu e_R^i + \frac{2}{3} (\bar{u}_L^i \gamma_\mu u_L^i + \bar{u}_R^i \gamma_\mu u_R^i) - \\ &- \frac{1}{3} (\bar{d}_L^i \gamma_\mu d_L^i + \bar{d}_R^i \gamma_\mu d_R^i) . \end{aligned} \quad (21.13)$$

This can happen only if  $e = g_1 \cos \theta_W = g_2 \sin \theta_W$ . It follows that:

$$\tan \theta_W = \frac{g_1}{g_2} , \quad e = \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} . \quad (21.14)$$

Inserting this into (21.12) we learn that the coupling of the  $Z$ -boson is:

$$\frac{1}{\sqrt{g_1^2 + g_2^2}} \left( -\frac{1}{2} g_1^2 J^{\mu Y} + g_2^2 J^{\mu 3} \right) Z_\mu \equiv g_z J^{\mu\text{NC}} Z_\mu . \quad (21.15)$$

Thus the weak neutral current is given by:

$$g_z J_\mu^{\text{NC}} = \frac{1}{\sqrt{g_1^2 + g_2^2}} \left( -\frac{1}{2} g_1^2 J_\mu^Y + g_2^2 J_\mu^3 \right) , \quad (21.16)$$

where  $g_z$  is the coupling constant we associate to the  $Z$ -boson. This is a convention, because only the combination  $g_z J_\mu^{\text{NC}}$  appears in formulae. For convenience we choose:  $g_z = g_2 / \cos \theta_W = \sqrt{g_1^2 + g_2^2}$ . With this, the weak neutral current is:

$$\begin{aligned} J_\mu^Z &= J_\mu^{\text{NC}} = -\frac{1}{2} \frac{g_1^2}{g_1^2 + g_2^2} J_\mu^Y + \frac{g_2^2}{g_1^2 + g_2^2} J_\mu^3 = \\ &= -\frac{1}{2} \sin^2 \theta_W J_\mu^Y + \cos^2 \theta_W J_\mu^3 = J_\mu^3 - \sin^2 \theta_W J_\mu^{\text{em}} , \end{aligned} \quad (21.17)$$

where we have written two different forms that are both useful.

Taking a look at the first of (21.10) show us that the charged currents  $J^{\mu\pm}$  involve only the left chiral fermions and have the so called V(ector)–A(xial vector) structure.  $J_\mu^{\text{em}}$  given by (21.13) has pure vector nature. Eqs. (21.10) and (21.17) clearly show that, unlike the  $W^\pm$  bosons, the  $Z$ -boson does not have  $V - A$  couplings with the fermions. It must be kept in mind that when coupling it to  $Z_\mu$ , this current should be multiplied by  $g_z = g_2 / \cos \theta_W$ . Note that the expression of the current will remain the same even when it is written in terms of the mass eigenstates  $d^i$  of instead of  $d^i$ .

The weak NC can also be written in terms of the  $T_3$  and  $Y$  of the various fermions and also as a combination of  $V$  and  $A$  currents as follows.

$$J_\mu^Z = \sum_f \left[ \bar{f} \gamma_\mu f_L g_L^f + \bar{f} \gamma_\mu f_R g_R^f \right] = \left[ \frac{1}{2} g_V^f \bar{f} \gamma_\mu f - \frac{1}{2} g_A^f \bar{f} \gamma_\mu \gamma_5 f \right]. \quad (21.18)$$

Here the sum is over all fermions  $f^i = u^i, d^i, e^i, \nu^i$  ( $i = 1, 2, 3$ ). The couplings  $g_L^f, g_R^f, g_V^f$  and  $g_A^f$  can be read off from (21.10) and (21.17) to be

$$\begin{aligned} g_L^f &= T_3(f_L) - \sin^2 \theta_W Q_f, & g_V^f &= T_3(f_L) + T_3(f_R) - 2 Q_f \sin^2 \theta_W, \\ g_R^f &= T_3(f_R) - \sin^2 \theta_W Q_f, & g_A^f &= T_3(f_L) - T_3(f_R). \end{aligned} \quad (21.19)$$

In the above equation, we have written down  $T_3(f_R)$  explicitly, which in the EW model is zero, with a view to generalize the expressions for the weak neutral current, should the fermions belong to other representations of  $SU(2)_L \times U(1)_Y$ , other than the one in the EW model. Recall that  $Q_f$  is the electromagnetic charge of the fermion in units of positron charge. The values of  $g_A^f, g_V^f, g_L^f$  and  $g_R^f$  for the fermions of the EW model are given in the table:

$f$	$\nu$	$e^-$	$u$	$d$
$g_L^f$	$\frac{1}{2}$	$-\frac{1}{2} + \sin^2 \theta_W$	$\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W$	$-\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W$
$g_R^f$	0	$\sin^2 \theta_W$	$-\frac{2}{3} \sin^2 \theta_W$	$\frac{1}{3} \sin^2 \theta_W$
$g_A^f$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$g_V^f$	$\frac{1}{2}$	$-\frac{1}{2} + 2 \sin^2 \theta_W$	$\frac{1}{2} - \frac{4}{3} \sin^2 \theta_W$	$-\frac{1}{2} + \frac{2}{3} \sin^2 \theta_W$

## 21.5 Strengths of the CC and NCs

Note now that the form for the NC of (21.18) is exactly the same, for all the fermions of a given electrical charge and given values of the  $SU(2)_L$  quantum

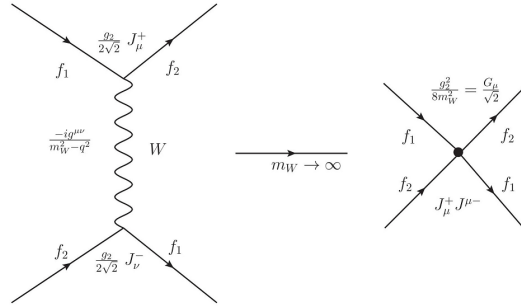
numbers. Since in the EW model, all the quarks or leptons of a given electric charge and handedness belong to the same representation of  $SU(2)$  the weak NC automatically conserves 'flavor', be it the leptonic one or the quark one. This is indeed quite reassuring since the experiments had shown that while 'flavor' changing charged weak current (21.10) exist, decays caused by 'flavor' changing weak NC, which are either forbidden or suppressed by orders of magnitude. Their absence at the tree level is automatically guaranteed in the EW model, just by the particle content.

We see that in the EW model, the weak NC couplings are completely determined by  $g_2$  and  $\sin\theta_W$ . The weak NC involving  $\nu^i$  is pure left handed just like the corresponding charged current, where (as for the charged fermions) the  $V - A$  mixture depends on the electromagnetic charge of the fermion because the relative weight of  $L$  and  $R$  currents is decided by the hypercharge  $Y$ .

While the strength of the axial current is completely decided by the  $T_3$  value of  $f_L^i$ , the vector coupling depends on the weak mixing angle  $\theta_W$ . The experimentally determined value of  $\sin^2\theta_W \sim 0.23$ . As a result the weak NC coupling of the charged lepton ( $e, \mu, \tau$ ) is in fact close to zero.

The interaction of all the quarks and leptons with the EW gauge bosons is encoded in the currents  $J_\mu^{\text{em}}$ ,  $J_\mu^\pm$  and  $J_\mu^Z$  given by (21.13), first of (21.10) and (21.18). In low energy reactions, the appropriate way to adjudge the strength of processes mediated by the weak NC is to derive the current-current form of the interaction Lagrangian starting from (21.18). This is done by considering the matrix element of a four fermion scattering process and taking the limit in which the mass of the exchanged gauge boson is infinite.

Let us consider the scattering process  $f_1 + f_2 \rightarrow f_1 + f_2$ , through the exchange of a massive  $W^\pm$  (i.e. via CC), as indicated in the figure below.



The effective current-current Lagrangian for this scattering process can be written as:

$$\mathcal{L}_{\text{eff}}^{\text{CC}} = -\frac{g_2^2}{8M_W^2} J_\mu^+ J^{-\mu} = -\frac{G_\mu}{\sqrt{2}} J_\mu^+ J^{-\mu} , \tag{21.20}$$

with  $J_\mu^\pm$  as given by (21.10). On comparing with the current-current interactions of the pre gauge theory days, one then gets:

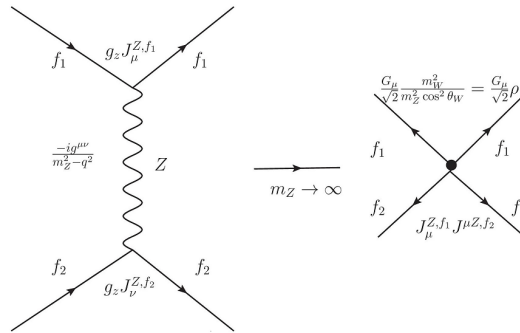
$$\frac{G_\mu}{\sqrt{2}} = \frac{g_2^2}{8M_W^2} = \frac{e^2}{8M_W^2 \sin^2 \theta_W} , \tag{21.21}$$

where

$$G_\mu V_{ud} = G_F . \tag{21.22}$$

It can be noted here that since  $|\sin \theta_W| < 1$ , the experimentally measured value of  $G_\mu$  and  $e$ , tells us that  $M_W > 37.43 \text{ GeV}$ . For the limiting value of  $\sin \theta_W \sim 1$  we get  $M_W \sim 100 \text{ GeV}$ .

One can similarly write down the effective NC interaction effective Lagrangian under the approximation that the Z boson mass is large, by considering the four-fermion scattering process shown in the Figure below.



This is given by

$$\mathcal{L}_{\text{eff}}^{\text{NC}} = -\frac{g_z^2}{2} \left( \sum_f J_\mu^{Z,f} \right) \left( \sum_f J^{\mu,Z,f} \right). \quad (21.23)$$

If one calculates the matrix elements for scattering process (20.14) taking place via the interaction of (21.20) and (21.23) respectively,  $\mathcal{M}_{CC}$  and  $\mathcal{M}_{NC}$ , it can be seen that their ratio is given in terms of  $M_Z$ ,  $M_W$  and  $\sin\theta_W$  as:

$$\frac{\mathcal{M}_{NC}}{\mathcal{M}_{CC}} = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} \equiv \rho. \quad (21.24)$$

This effective Lagrangian involves couplings  $g_2$ ,  $g_1$  and  $M_W$ ,  $M_Z$ . We can use the two measured couplings  $G_\mu$  and  $\alpha_{\text{em}}$  along with  $\rho$  and one arbitrary parameter of the model the weak mixing angle  $\sin\theta_W$ .  $M_W$ ,  $M_Z$  are then given in terms of these and we have traded  $g_1$ ,  $g_2$  for  $G_\mu$  and  $\alpha_{\text{em}}$ .

Note also that in these discussions we have completely sidestepped the issue of how the non-zero masses for the gauge bosons and the fermions written can be made consistent with gauge invariance. In case of the gauge bosons the loss of gauge invariance also means loss of renormalizability and hence consequently of the ability to make any predictions. So one of the problems to be addressed is how to generate the mass terms below in a gauge invariant manner.

$$\mathcal{L}_{\text{mass}} = \frac{1}{2} M_Z^2 Z_\mu Z^\mu + M_W^2 W_\mu^+ W^{-\mu} + \sum_i m_i [\bar{\psi}_{iL} \psi_{iR} + \bar{\psi}_{iR} \psi_{iL}]. \quad (21.25)$$

It should be noted that the sum in (21.25) is over all the fermions except the neutrinos which are assumed to be massless here in this discussion.

**Exercise 21.1:** Assume a universal  $V-A$  interaction, compute (or estimate) the ratio of rates:  $\Gamma(\pi^- \rightarrow \mu^- \bar{\nu})/\Gamma(\pi^- \rightarrow e^- \bar{\nu})$ . How would this ratio change if the universal weak interaction coupling were scalar? Pseudoscalar?

**Exercise 21.2:** Check the unitarity condition for the Kobaiashi-Maskava matrix. Which elements are restricted due to unitarity requirement?

**Exercise 21.3:** Find the explicit form for the spinor left-right projection operators:  $P_L^2$ ,  $P_R^2$ ,  $P_L P_R$ ,  $P_L + P_R$ .

**Exercise 21.4:** Write out covariant derivatives for the left and right leptons.

## Part VIII

# Lecture – Spontaneous Symmetry Breaking





## Chapter 22

# The Mechanism of SSB

*Spontaneous Symmetry Breaking* (SSB) refers to the fact that the lowest energy state, the vacuum, may not be invariant under all symmetries of the theory, in other words, several vacua are possible. In this lecture we shall consider SSB and corresponding mass generation mechanism, which play significant roles in both particle physics and condensed matter physics.

### 22.1 The Beginnings of SSB

Historically, the concept of SSB first emerged in condensed matter physics. The prototype case is the 1928 Heisenberg theory of the ferromagnet as an infinite array of spin  $1/2$  magnetic dipoles, with spin-spin interactions between nearest neighbors such that neighboring dipoles tend to align. Although the theory is rotationally invariant, below the critical Curie temperature  $T_c$  the actual ground state of the ferromagnet has the spin all aligned in some particular direction (i.e. a magnetization pointing in that direction), thus not respecting the rotational symmetry. What happens is that below  $T_c$  there exists an infinitely degenerate set of ground states, in each of which the spins are all aligned in a given direction. A complete set of quantum states can be built upon each ground state. We thus have many different "possible worlds" (sets of solutions to the same equations), each one built on one of the possible orthogonal (in the infinite volume limit) ground states.

The same picture can be generalized to QFT, the ground state becoming

the vacuum state. This means that there may exist symmetries of the laws of nature which are not manifest to us because the physical world in which we live is built on a vacuum state which is not invariant under them. In other words, the physical world of our experience can appear to us very asymmetric, but this does not necessarily mean that this asymmetry belongs to the fundamental laws of nature. SSB offers a key for understanding (and utilizing) this physical possibility.

The concept of SSB was transferred from condensed matter physics to QFT in the early 1960s, thanks especially to works by Nambu and Jona-Lasinio. The idea of SSB was introduced and formalized in particle physics on the grounds of an analogy with the breaking of (electromagnetic) gauge symmetry in the 1957 theory of superconductivity by Bardeen, Cooper and Schrieffer. The application of SSB to particle physics in the 1960s and successive years led to profound physical consequences and played a fundamental role in the edification of the SM.

Let us mention the main results that obtain in the case of the spontaneous breaking of a continuous internal symmetry in QFT:

- *Goldstone theorem.* In the case of a global continuous symmetry, massless bosons (known as "Goldstone bosons") appear with the spontaneous breakdown of the symmetry according to a theorem first stated by Goldstone in 1960. The presence of these massless bosons, first seen as a serious problem since no particles of the sort had been observed in the context considered, was in fact the basis for the solution – by means of the so-called Higgs mechanism (see the next point) – of another similar problem, that is the fact that the 1954 Yang-Mills theory of non-Abelian gauge fields predicted unobservable massless particles, the gauge bosons;
- *Higgs mechanism.* According to a "mechanism" established in a general way in 1964 independently by (i) Higgs, (ii) Brout and Englert, (iii) Guralnik, Hagen and Kibble, in the case that the internal symmetry is promoted to a local one, the Goldstone bosons "disappear" and the gauge bosons acquire a mass. The Goldstone bosons are "eaten up" to give mass to the gauge bosons, and this happens without (explicitly) breaking the gauge invariance of the theory.

Note that this mechanism for the mass generation for the gauge fields is also what ensures the renormalizability of theories involving massive

gauge fields (such as the Glashow-Weinberg-Salam EW theory developed in the second half of the 1960s), as first generally demonstrated by Veltman and 't Hooft in the early 1970s;

- *Dynamical symmetry breaking.* In such theories as the unified model of EW interactions, the SSB responsible (via the Higgs mechanism) for the masses of the gauge vector bosons is because of the symmetry-violating vacuum expectation values of scalar fields (the so-called Higgs fields) introduced *ad hoc* in the theory.

For different reasons (first of all, the initially *ad hoc* character of these scalar fields for which there was no experimental evidence until the results obtained in 2012 at the LHC) some attention has been drawn to the possibility that the Higgs fields could be phenomenological rather than fundamental, that is bound states resulting from a specified dynamical mechanism.

SSB allows symmetric theories to describe asymmetric reality. In short, SSB provides a way of understanding the complexity of nature without renouncing fundamental symmetries. But why should we prefer symmetric to asymmetric fundamental laws? In other words, why assume that an observed asymmetry requires a cause, which can be an explicit breaking of the symmetry of the laws, asymmetric initial conditions, or SSB?

Note that this assumption is very similar to the one expressed by Curie in his famous 1894 paper. Curie's principle: *the symmetries of the causes must be found in the effects* (or equivalently, the asymmetries of the effects must be found in the causes), when extended to include the case of SSB, is equivalent to a methodological principle according to which an asymmetry of the phenomena must come from the breaking (explicit or spontaneous) of the symmetry of the fundamental laws. What the real nature of this principle is remains an open issue.

Finally, let us mention the argument that is sometimes made in the literature that SSB implies that Curie's principle is violated because a symmetry is broken "spontaneously", that is without the presence of any asymmetric cause. Now it is true that SSB indicates a situation where solutions exist that are not invariant under the symmetry of the law (dynamical equation) without any explicit breaking of this symmetry. But, as we will see, the symmetry of the "cause" is not lost, it is conserved in the ensemble of the solutions (the whole "effect").

## 22.2 Breaking Global $U(1)$

Before studying the details of the SSB let us consider some examples of SSB in field theory.

Suppose we have a complex scalar bosons,  $\phi$  and  $\phi^\dagger$ , with the Lagrangian,

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi^\dagger\partial_\mu\phi - \frac{1}{2}m^2\phi^\dagger\phi. \quad (22.1)$$

Naturally we can write this as

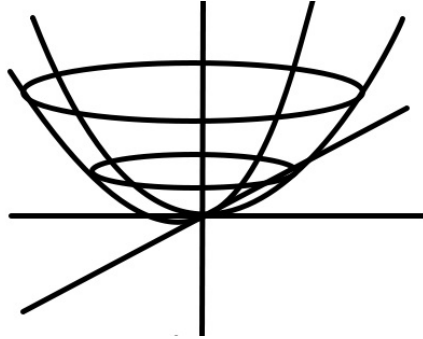
$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi^\dagger\partial_\mu\phi - V(\phi^\dagger, \phi), \quad (22.2)$$

where

$$V(\phi^\dagger, \phi) = \frac{1}{2}m^2\phi^\dagger\phi. \quad (22.3)$$

This Lagrangian has the  $U(1)$  symmetry we discussed in 19.1.

We can graph  $V(\phi^\dagger, \phi)$ , plotting  $V$  vs.  $|\phi|$ ,

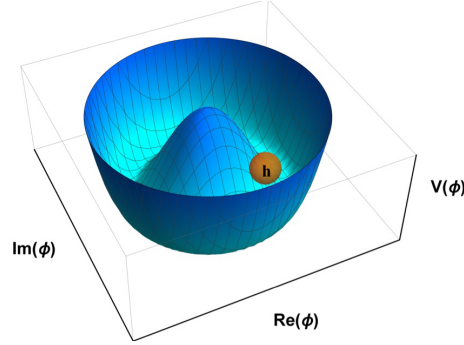


We see a 'bowl' with  $V_{min}$  at  $|\phi|^2 = 0$ . The vacuum of any theory ends up being at the lowest potential point, and therefore the vacuum of this theory is at  $\phi = 0$ , as we would expect.

Now, let's change the potential. Consider

$$V(\phi^\dagger, \phi) = \frac{1}{2}\lambda m^2(\phi^\dagger\phi - \Phi^2)^2, \quad (22.4)$$

where  $\lambda$  and  $\Phi$  are real constants. Notice that the Lagrangian will still have the global  $U(1)$  symmetry from before. But, now if we graph  $V(\phi)$  in the  $\text{Re}(\phi)$  -  $\text{Im}(\phi)$  plain, we get,



where now the vacuum  $V_{min}$  is represented by the circle at  $|\phi| = \Phi$ . In other words, there are an infinite number of vacuums in this theory. Because the circle drawn in the figure above represents a rotation through field space, this degenerate vacuum is parameterized by  $e^{i\alpha}$ , the global  $U(1)$ . There will be a vacuum for every value of  $\alpha$ , located at  $|\phi| = \Phi$ .

In order to make sense of this theory, we must choose a vacuum by hand. Because the theory is completely invariant under the choice of the  $U(1)$ -phase in  $e^{i\alpha}$ , we can choose any  $\alpha$  and define that as our true vacuum. So, we choose  $\alpha$  to make our vacuum at  $\phi = \Phi$ , or where  $\phi$  is real and equal to  $\Phi$ . We have thus, in a sense, *Gauged Fixed* the symmetry in the Lagrangian, and the  $U(1)$  symmetry is no longer manifest.

Now we need to rewrite this theory in terms of our new vacuum. We therefore expand around the constant vacuum value  $\Phi$  to have the new field

$$\phi \equiv \Phi + h + i\beta, \quad (22.5)$$

where  $h$  and  $\beta$  are new real scalar fields (so  $\phi^\dagger = \Phi + h - i\beta$ ). We can now write out the Lagrangian as

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2}\partial^\mu(h - i\beta)\partial_\mu(h + i\beta) - \\ &\quad -\frac{1}{2}\lambda m^2 [(\Phi + h - i\beta)(\Phi + h + i\beta) - \Phi^2]^2 = \\ &= -\frac{1}{2}\partial^\mu h\partial_\mu h - \frac{1}{2}4\lambda m^2\Phi^2 h^2 - \frac{1}{2}\partial^\mu\beta\partial_\mu\beta - \\ &\quad -\frac{1}{2}\lambda m^2 (4\Phi h^3 + 4\Phi h\beta^2 + h^4 + h^2\beta^2 + \beta^4). \end{aligned} \quad (22.6)$$

This is now a theory of a massive real scalar field  $h$  (with mass  $= \sqrt{4\lambda m^2\Phi^2}$ ), a massless real scalar field  $\beta$ , and five different types of interactions (one

allowing three  $h$ 's to interact, the second allowing one  $h$  and two  $\beta$ 's, the third allowing four  $h$ 's, the fourth allowing two  $h$ 's and two  $\beta$ 's, and the last allowing four  $\beta$ 's). In other words, there are five different types of vertices allowed in the Feynman diagrams for this theory.

Furthermore, notice that this theory has no obvious  $U(1)$  symmetry. For this reason, writing the field in terms of fluctuations around the vacuum we choose is called 'breaking' the symmetry. The symmetry is still there, but it can't be seen in this form.

Finally, notice that breaking the symmetry has resulted in the addition of the massless field  $\beta$ . It turns out that breaking global symmetries as we have done always results in a massless boson. Such particles are called *Goldstone Bosons*.

### 22.3 Breaking Local $U(1)$

In the previous section, we broke a global  $U(1)$  symmetry. In this section, we will break a local  $U(1)$  and see what happens. We begin with the Lagrangian for a complex scalar field with a gauged  $U(1)$ :

$$\mathcal{L} = -\frac{1}{2} \left[ (\partial^\mu - iqA^\mu) \phi^\dagger (\partial_\mu + iqA_\mu) \phi \right] - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - V(\phi^\dagger, \phi), \quad (22.7)$$

where we have taken the external source  $J^\mu = 0$ . Let's once again assume  $V(\phi^\dagger, \phi)$  has the form of equation (22.4), so the vacuum has the  $U(1)$  degeneracy at  $|\phi| = \Phi$ .

Because our  $U(1)$  is now local, we choose  $\alpha(x)$  so that not only is the vacuum real, but also so that  $\phi$  is always real. We therefore expand

$$\phi = \Phi + h, \quad (22.8)$$

where  $h$  is a real scalar field representing fluctuations around the vacuum we chose. Now,

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2} [(\partial^\mu - iqA^\mu) (\Phi + h) (\partial_\mu + iqA_\mu) (\Phi + h)] - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \\ &\quad - \frac{1}{2} \lambda m^2 [(\Phi + h) (\Phi + h) - \Phi^2]^2 = \quad (22.9) \\ &= -\frac{1}{2} \partial^\mu h \partial_\mu h - \frac{1}{2} 4\lambda m^2 \Phi^2 h^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} - \frac{1}{2} q^2 \Phi^2 A^2 + \mathcal{L}_{int}, \end{aligned}$$

where the allowed interaction terms in  $\mathcal{L}_{int}$  include a vertex connecting an  $h$  and two  $A^\mu$ 's, four  $h$ 's, and three  $h$ 's.

Before breaking we had a complex scalar field  $\phi$  and a massless vector field  $A^\mu$  with two polarization states (because it is a photon). Now, we have a single real scalar  $h$  with mass  $= \sqrt{4\lambda m^2 \Phi^2}$  and a field  $A^\mu$  with mass  $= q\Phi$ . In other words, our force-carrying particle  $A^\mu$  has gained mass! We started with a theory with no mass, and by merely breaking the symmetry, we have introduced mass into our theory.

This mechanism for introducing mass into a theory, called the *Higgs Mechanism*, was first discovered by Higgs in 1960s, and the resulting field  $h$  is called the *Higgs Boson*.

So, whereas the consequence of global symmetry breaking is a massless boson called a Goldstone boson, the consequence of a local symmetry breaking is that the gauge field, which came about as a result of the symmetry being local, acquires mass.

## 22.4 SSB in Non-Abelian Gauge Theory

As we already know, given a field transforming in a particular representation  $R$ , the gauge fields  $A^\mu$  will be  $D(R) \times D(R)$  matrices.

Once we know what representation we are working in, and therefore know the generators  $T_R^a$ , it turns out that it is always possible to write the gauge fields in terms of the generators. We can think of the generators as basis vectors which span the parameter space for the group. Because the gauge fields live in the  $N \times N$  space as well, we can write them in terms of the generators. That is, instead of the gauge fields being  $N \times N$  matrices on their own, we will use the  $N \times N$  matrix generators as basis vectors, and then the gauge fields can be written as scalar coefficients of each generator:

$$A^\mu = A_a^\mu T_R^a, \quad (22.10)$$

where  $a$  is understood to be summed, and each  $A_a^\mu$  is now a scalar function rather than a  $D(R) \times D(R)$  matrix (the advantage of this is that we can continue to think of the gauge fields as scalars with an extra index, rather than as matrices).

As a note, we haven't done anything particularly profound here. We are merely writing each component of the  $D(R) \times D(R)$  matrix  $A^\mu$  in terms of the  $D(R) \times D(R)$  generators, allowing us to work with a scalar field  $A_a^\mu$  rather than the matrix field  $A^\mu$ .

We now actually view each  $A_a^\mu$  as a separate field. So, if a group has  $N$  generators, we say there are  $N$  gauge fields associated with it, each one having 4 spacetime components  $\mu$ . In matrix components, we will have

$$(A^\mu)_{ij} = (A_a^\mu T_R^a)_{ij} . \quad (22.11)$$

Then, the covariant derivative will be

$$(D_\mu \phi)_j = \partial_\mu \phi_j(x) - ig[A_{\mu a}(x) T_R^a]_{jk} \phi_k(x) . \quad (22.12)$$

We may assume that the field strength  $F^{\mu\nu}$  can also be expressed in terms of the generators, so that we have

$$F^{\mu\nu} = F_a^{\mu\nu} T^a , \quad (22.13)$$

or

$$(F^{\mu\nu})_{ij} = (F_a^{\mu\nu} T^a)_{ij} . \quad (22.14)$$

Now, we can write the Lagrangian of non-Abelian vector field in terms of the new basis:

$$\begin{aligned} \mathcal{L}_{Kin} = -\frac{1}{2} \text{Tr}(F_{\mu\nu} F^{\mu\nu}) &= -\frac{1}{2} \text{Tr}(F_a^{\mu\nu} T^a F_{\mu\nu b} T^b) = \\ &= -\frac{1}{2} F_a^{\mu\nu} F_{\mu\nu b} \text{Tr}(T^a T^b) = \\ &= -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a . \end{aligned} \quad (22.15)$$

We have raised the index  $a$  on the second field strength term simply to explicitly imply the summation over it. The fact that it is raised doesn't change its value in this case; it is merely notational.

Furthermore, we can invert the expression (22.13):

$$\begin{aligned} F^{\mu\nu} = F_a^{\mu\nu} T^a &\Rightarrow F^{\mu\nu} T^b = F_a^{\mu\nu} T^a T^b , \\ &\Rightarrow \text{Tr}(F^{\mu\nu} T^b) = F_a^{\mu\nu} \text{Tr}(T^a T^b) = \frac{1}{2} F_b^{\mu\nu} , \\ &\Rightarrow F_a^{\mu\nu} = 2 \text{Tr}(F^{\mu\nu} T^a) . \end{aligned} \quad (22.16)$$



Above we broke the  $U(1)$  symmetry, which only had one generator. However, if we break larger groups we may only break part of it. For example, we will see that  $SU(3)$  has an  $SU(2)$  subgroup. It is actually possible to break only the  $SU(2)$  part of the  $SU(3)$ . So, three of the  $SU(3)$  generators are broken (the three corresponding to the  $SU(2)$  subgroup/subalgebra), and the other five are unbroken. Because we are now writing our gauge fields using the generators as a basis, this means that three of the gauge fields are broken, while five of the gauge fields are not.

Finally, recall from Sec. 22.3 that breaking a local symmetry results in a gauge field gaining mass. We seek now to elucidate the relationship between breaking a symmetry and a field gaining mass. First, we can summarize as follows: Gauge fields corresponding to broken generators get mass, while those corresponding to unbroken generators do not. The unbroken generators form a new gauge group that is smaller than the original group that was broken.

In Sec. 22.3, we saw that breaking a symmetry gave the gauge field mass. We can note also the opposite effect – giving a gauge field mass will break the symmetry.

To make this clearer, we begin with a very simple example, then move on to a more complicated example.

## 22.5 Simple Examples of SSB

Consider a theory with three real massless scalar fields  $\phi_i$  ( $i = 1, 2, 3$ ) and with Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i\partial_\mu\phi^i, \quad (22.17)$$

which is clearly invariant under the global  $SO(3)$  rotation

$$\phi_i \rightarrow R^{ij}\phi_j, \quad (22.18)$$

where  $R^{ij}$  is an element of  $SO(3)$ , because the Lagrangian is a dot product in field space, and we know that dot products are invariant under  $SO(3)$ .

Now, let's say that one of the fields, say  $\phi_1$ , gains mass. The new Lagrangian will then be

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i\partial_\mu\phi^i - \frac{1}{2}m^2\phi_1^2. \quad (22.19)$$

So this Lagrangian is no longer invariant under the full  $SO(3)$  group, which mixes any two of the three fields. Rather, it is only invariant under rotations in field space that mix  $\phi_2$  and  $\phi_3$  or  $SO(2)$ . In other words, giving one field mass broke  $SO(3)$  to the smaller  $SO(2)$ .

As another simple example consider the model with five massless complex scalar fields  $\phi_i$ , with the Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i^\dagger\partial_\mu\phi^i . \quad (22.20)$$

This will be invariant under any  $SU(5)$  transformation. Then let's say we give two of the fields,  $\phi_1$  and  $\phi_2$  (equal) mass. The new Lagrangian will be

$$\mathcal{L} = -\frac{1}{2}\partial^\mu\phi_i^\dagger\partial_\mu\phi^i - \frac{1}{2}m(\phi_1^\dagger\phi_1 + \phi_2^\dagger\phi_2) . \quad (22.21)$$

So now, we no longer have the full  $SU(5)$  symmetry, but we do have the special unitary transformations mixing  $\phi_3, \phi_4,$  and  $\phi_5$ . This is an  $SU(3)$  subgroup. Furthermore, we can do a special unitary transformation mixing  $\phi_1$  and  $\phi_2$ . This is an  $SU(2)$  subgroup. So, we have broken

$$SU(5) \rightarrow SU(3) \otimes SU(2) . \quad (22.22)$$

Before considering a more complicated example of this, we further discuss the connection between symmetry breaking and fields gaining mass.

When we introduced SSB, recall that we shifted the potential minimum from  $V_{min}$  at  $\phi = 0$  to  $V_{min}$  at  $|\phi| = \Phi$ . But we were discussing this in very classical language. We can interpret all of this in a more 'quantum' way in terms of *Vacuum Expectation Values* (VEV). As we said, the vacuum of a theory is defined as the minimum potential field configuration. For the  $V_{min}$  at  $\phi = 0$  potential, the VEV of the field  $\phi$  was at 0, or

$$\langle 0|\phi|0\rangle = 0 . \quad (22.23)$$

However, for the  $V_{min}$  at  $|\phi| = \Phi$  potential, we have

$$\langle 0|\phi|0\rangle = \Phi . \quad (22.24)$$

So, in quantum mechanical language, symmetry breaking occurs when a field, or some components of a field, take on a non-zero VEV.

This seems to be what is happening in nature. At higher energies, there is some *Master Theory* with some gauge group defining the physics, and all of the fields involved have zero VEV's. At lower energies, for whatever reason (the reason for this is not well understood at the time of this writing), some of the fields take on non-zero VEV's, which break the symmetry into smaller groups, giving mass to certain fields through the Higgs Mechanism discussed in Sec. 22.3. We call the theory with the unbroken gauge symmetry at higher energies the more fundamental theory (analogous to equation (22.2)), and the Lagrangian which results from breaking the symmetry (analogous to (22.6)) the *Low Energy Effective Theory*.

This is how mass is introduced into the SM. It turns out that if a theory is renormalizable one can prove that any lower energy effective theory that results from breaking the original theory's symmetry is also renormalizable, even if it doesn't appear to be. Because the actions that appear to describe the universe at the energy level we live at (and the levels attainable by current experiment) are not renormalizable when they have mass terms. We work with a larger theory which has no massive particles but can be renormalized, and use the Higgs Mechanism to give various particles mass. So, whereas the physics we see at low energies may not appear renormalizable, if we can find a renormalizable theory which breaks down to our physics, we are safe.

Now, we consider a slightly more complicated (and realistic) example of symmetry breaking.

## 22.6 Breaking of $SU(N)$

Consider the gauge group  $SU(N)$ , acting on  $N$  complex scalar fields  $\phi_i$  ( $i = 1, \dots, N$ ) in the fundamental representation  $\mathbf{N}$ . As in Sec. 22.3, where we made use of the  $U(1)$  symmetry to make the VEV real, we can use  $SU(N)$  to not only make the VEV real, but also to rotate it to a single component of the field,  $\phi_N$ ,

$$\begin{aligned} \langle 0 | \phi_i | 0 \rangle &= 0, & (i = 1, \dots, N-1) \\ \langle 0 | \phi_N | 0 \rangle &= \Phi. \end{aligned} \tag{22.25}$$

So, we expand  $\phi_N$  around this new vacuum:

$$\begin{aligned} \phi_i &= \phi_i, & (i = 1, \dots, N-1) \\ \phi_N &= \Phi + \chi. \end{aligned} \tag{22.26}$$

This means that, in the vacuum configuration, the fields will have the form

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix}_{vac} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \Phi \end{pmatrix}. \quad (22.27)$$

How will the action of  $SU(N)$  be affected by this VEV? If we consider a general element of  $SU(N)$  acting on this,

$$\begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1N} \\ U_{21} & U_{22} & \cdots & U_{2N} \\ \vdots & \vdots & \ddots & \cdots \\ U_{N1} & U_{N2} & \vdots & U_{NN} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \Phi \end{pmatrix} = \begin{pmatrix} U_{1N} \\ U_{2N} \\ \vdots \\ U_{NN} \end{pmatrix}. \quad (22.28)$$

So, only elements of  $SU(N)$  with non-zero elements in the last column will be affected by this VEV. But the other  $N - 1$  elements' rows and columns are unaffected. This means that we have an  $SU(N - 1)$  symmetry left. Or in other words, we have broken  $SU(N) \rightarrow SU(N - 1)$  with this VEV.

Let's consider a specific example of the  $SU(3)$  case. Notice that exactly three generators of this group have all zeros in the last column:  $\lambda^1$ ,  $\lambda^2$ , and  $\lambda^3$ . We expect these three to give an  $SU(2)$  subgroup. Looking at the upper left  $2 \times 2$  boxes in those three generators, we can see that they are the Pauli matrices, the generators of  $SU(2)$ . So, if we give a non-zero VEV to the fields transforming under  $SU(3)$ , we see that they do indeed break the  $SU(3)$  to  $SU(2)$ . The other five generators of  $SU(3)$  will be affected by the VEV, and consequently the corresponding fields will acquire mass.

**Exercise 22.1:** Why within the SSB mechanism with complex scalar fields VEVs are assumed to be real parameters?

**Exercise 22.2:** Suppose we add a cubic term,  $\sim \phi^3$  to the Higgs potential  $U(\phi)$ . Show that the degeneracy in the minimum of  $U(\phi)$  is now removed. Find the true minimum of the potential. Also, show that, as a function of the parameter in front of cubic term, the VEV changes discontinuously from -VEV to +VEV as this parameter changes from positive to negative values going through 0.

**Exercise 22.3:** Explain why the Higgs potential can contain terms with only even powers of the field.

## Chapter 23

# SSB in Condensed Matter

As we mentioned in previous chapter, first examples of SSB appeared in condensed matter physics, and the idea was transplanted into particle physics by analogy. The simple forms of SSB is realized in condensed matter theory. In this Chapter we shall briefly mention several such models. We shall largely restrict ourselves to continuous transitions for which relevant experiments have been performed.

### 23.1 Ginzburg-Landau Model of Superconductivity

In the Ginzburg-Landau Model for low- $T_c$  superconductors the scalar effective order parameter field  $\phi$  is derived from the  $L = S = 0$  Cooper pairs of electrons (with momenta close to the Fermi surface) as

$$\phi \sim \langle \psi_{\downarrow} \psi_{\uparrow} \rangle_{L=S=0} . \quad (23.1)$$

This is an Abelian Higgs model with the free energy in the absence of external fields:

$$F = \int d^3x \left[ \frac{\hbar^2}{2m^*} \left| \left( -i\nabla - \frac{e^*}{\hbar c} \mathbf{A} \right) \phi \right|^2 + \lambda (|\phi|^2 - \eta^2)^2 \right] , \quad (23.2)$$

where  $e^* = 2e$ ,  $m^* = 2m_e$  characterize the Cooper pair. The Meissner effect, whereby the magnetic field penetrates into the bulk superconductor, has its penetration length determined by the mass of the vector field in the broken

phase. Whether this is a Type-I or Type-II superconductor depends on the ratio  $\kappa$  of this length to the coherence length (the London length) of the order parameter  $\phi$ , determined by the Higgs mass. Large  $\kappa$  is Type-II, small  $\kappa$  is Type-I.

## 23.2 High- $T_c$ Superconductors

High- $T_c$  superconductivity is a complicated phenomenon. For our purposes, we adopt an idealized explicitly broken  $SO(5)$  model in two dimensions. The basic idea is that doping an antiferromagnetic leads to d-wave superconductivity. The bosonic effective order parameters are, again, constructed from fermionic bilinears.

We begin with a global  $O(3)_{AF}$  L $\sigma$ M for anti-ferromagnetism with order parameter field (staggered magnetism)  $\vec{n}$  with potential

$$V(\vec{n}) = \frac{\lambda}{4}(\vec{n}^2 - 1)^2 . \quad (23.3)$$

The  $O(3)_{AF}$  symmetry is broken to  $O(2)$  by any ground state in the manifold  $\mathcal{M} = S^2$ . We can work equivalently with a NL $\sigma$ M, in which  $\vec{n}$  is constrained to  $|\vec{n}| = 1$ . This is extended to a five-component order parameter field  $\vec{N} = (\phi_1, n_1, n_2, n_3, \phi_2)$ , with  $O(5)$ -invariant potential

$$V(\vec{N}) = \frac{\lambda}{4}(\vec{N}^2 - 1)^2 . \quad (23.4)$$

We can work equivalently with a NL $\sigma$ M, in which  $|\vec{N}| = 1$ . Ultimately we shall couple  $\phi = (\phi_1 + i\phi_2)$  locally to the EM field as in Low- $T_c$  superconductors.

In the first instance  $\mathcal{M} = S^4$ . We break the  $O(5)$  invariance *explicitly* to  $O(3)_{AF} \times SO(2)$  by the addition of antiferromagnetic and doping terms to the potential, most simply as

$$V(\phi, \vec{n}) = \frac{1}{2}a_S|\phi|^2 + \frac{1}{2}a_A\vec{n}^2 + \frac{b}{4}(|\phi|^2 + \vec{n}^2)^2 , \quad (23.5)$$

where  $a_S, a_R < 0$ . We then couple  $\phi$  to EM as

$$F = \int d^2x \left[ \frac{\hbar^2}{2m^*} \left| \left( -i\nabla - \frac{e^*}{\hbar c} \mathbf{A} \right) \phi \right|^2 + \frac{\hbar^2}{2m^*} (\nabla \vec{n})^2 + V(\phi, \vec{n}) \right] . \quad (23.6)$$

Increasing the doping drives  $|a_S| > |a_R|$ , making the  $U(1) \sim SO(2)$  superconducting direction the global minimum. That is,

$$O(5) \rightarrow O(3)_{AF} \times SO(2) \rightarrow SO(2), \quad (23.7)$$

with  $\mathcal{M} = S^1$ . This  $U(1) \sim SO(2)$  is then broken as for the low-temperature superconductors. The nature of the transition is more complicated than the Type-I or Type-II options of the low-temperature case.

### 23.3 Superfluid ${}^3\text{He}$

${}^3\text{He}$  is a *Fermi Liquid*, which can become superfluid by the formation of p-wave ( $L = S = 1$ ) 'Cooper pairs' of  ${}^3\text{He}$  atoms.  $L$  and  $S$  are uncoupled at short distances, to give a global symmetry group

$$G = SO(3)_L \times SO(3)_S \times U(1)_N. \quad (23.8)$$

The effective order parameters form a  $3 \times 3$  matrix  $A_{\alpha i}(x)$ , formed from the Fermi bilinears  $\langle \psi(x)\psi(x) \rangle_{L=S=1}$ . The label  $i = 1, 2, 3$  is the orbital angular momentum label and  $a = 1, 2, 3$  the spin label. The  $U(1)_N$  describes the overall phase freedom. Above the transition all the elements of the matrix have zero values. Below the transition, some of these quantities become non-zero. The symmetry of the order parameter after the transition corresponds to the manifold of symmetries which remain unbroken.

The free energy of these states can be expressed in the framework of the phenomenological Ginzburg-Landau theory by a potential

$$\begin{aligned} V_{GL}(A) = & -\alpha A_{a,i}^* A_{a,i} + \beta_1 A_{a,i}^* A_{a,i}^* A_{b,j} A_{b,j} + \\ & + \beta_2 A_{a,i}^* A_{a,i} A_{b,j}^* A_{b,j} + \beta_3 A_{a,i}^* A_{b,i}^* A_{a,j} A_{b,j} + \\ & + \beta_4 A_{a,i}^* A_{b,i} A_{b,j}^* A_{a,j} + \beta_5 A_{a,i}^* A_{b,i} A_{b,j} A_{a,j}^*. \end{aligned} \quad (23.9)$$

The different possible symmetries of the order parameter  $A_{ai}$  are identified with local minima and saddle points in this 18-dimensional energy surface. There are two stable phases;

- The  $A$  phase, in which

$$SO(3)_S \times SO(3)_L \times U(1)_N \rightarrow SO(2)_{S_z} \times U(1)_{L_z - N/2} \times Z_2. \quad (23.10)$$

The manifold of ground states is

$$\mathcal{M}_A = G/H_A = S^2 \times SO(3)/Z_2 . \quad (23.11)$$

The order parameter in the  $A$  phase ground state is anisotropic in both spin and orbital spaces (the 'axial' state): most simply, it takes the form

$$A_{a,i}^0 = \Delta_A \hat{z}_a (\hat{x}_i + i\hat{y}_i) , \quad (23.12)$$

where  $\hat{x}, \hat{y}, \hat{z}$  are unit vectors in the  $x, y, z$  directions respectively.

- The  $B$  phase, in which the orbital and spin angular momenta are locked together as

$$SO(3)_S \times SO(3)_L \times U(1)_N \rightarrow SO(3)_{S+L} . \quad (23.13)$$

The manifold of ground states is now

$$\mathcal{M}_B = G/H_B = S^1 \times SO(3) . \quad (23.14)$$

As a result,  $A_{a,i}$  resembles a rotation matrix. Specifically, in the bulk  $B$  phase,  $A_{a,i}$  reduces to the arbitrary orthogonal rotation matrix  $R_{a,i}$ ,

$$A_{a,i} = \Delta R_{a,i} e^{i\Phi} . \quad (23.15)$$

The energy balance between the  $A$  and  $B$  phases is determined by the relation between the parameters  $\beta_i$ . At zero pressure, the  $B$  phase corresponds to the absolute minimum, while at pressures above 20 bar there is a temperature range in which the  $A$  phase is preferred.

## 23.4 Superfluid ${}^4\text{He}$ :

This is the global  $O(2)$  Goldstone model with the complex order parameter field

$$\phi = \rho e^{i\theta} . \quad (23.16)$$

In the 2-fluid model  $n_s = \rho^2$  is the superfluid density and

$$\mathbf{v}_s = \frac{\hbar}{m} \nabla \theta \quad (23.17)$$

the superfluid velocity. The Goldstone mode describes sound.



## 23.5 Other Systems

There are other systems whose transitions are understood well, which potentially have parallels with the early universe. In particular, we would cite:

- *Uniaxial nematic liquid crystals*, for which the order parameter in the nematic phase is the non-oriented director vector with ground state manifold  $RP^2 = S^2/Z_2$ . The transition is first order. However, when considering symmetry breaking at the interface of an isotropic-nematic transition the anchoring of the director at the interface forces it to lie on a cone, whereby the order parameter space is a circle  $S^1$ , corresponding to the familiar  $U(1)$  breaking.
- *Bose-Einstein condensates* (BEC), which allow for a great variety of symmetry breaking if species of atoms are mixed. For example, consider two-species BEC (two different ultra-cold atomic gases in a trap) with order parameter fields  $\phi_a(x)$  ( $a = 1, 2$ ). The trap-independent part of the potential can be written as

$$V = \lambda_1(|\phi_1|^2 - \eta_1^2)^2 + \lambda_2(|\phi_2|^2 - \eta_2^2)^2 + \beta|\phi_1|^2|\phi_2|^2, \quad (23.18)$$

with tunable parameters, which is no more than (23.5) rewritten for  $SO(4)$ . For a single species (1<sup>st</sup> term) we have an  $O(2)$  symmetry that is totally broken, as in  ${}^4\text{He}$ , and for two species an  $O(2) \times O(2)$  symmetry, broken to  $O(2)$ , with similarity to the breaking of the residual symmetry in high- $T_c$  superconductors.

In next section we shall consider the example of SSB in non-relativistic classical theory.

## 23.6 SSB in Classical Mechanics

The  $U(1)$  gauge field is usually induced from the gauge principle in a quantum theory. However, the gauge principle is alive also for the symmetry of translation in a non-relativistic classical theory and without any phase variable one can obtain the gauge field. Using this field theory it can be obtained the SSB and Higgs mechanism which is similar to the super-conductivity.

Let us start with  $N$  non-relativistic free particles in  $D$ -dimensional space with the Lagrangian:

$$L(\mathbf{z}(t), \dot{\mathbf{z}}(t)) = \frac{m}{2} \sum_{k=1}^N \dot{\mathbf{z}}_k^2 . \quad (23.19)$$

Then we can construct the conserved current field as

$$\mathbf{j}(x, t) = \sum_{k=1}^N \dot{\mathbf{z}}_k(t) \delta^D(x - z_k(t)) , \quad \rho(x, t) = \sum_{k=1}^N \delta^D(x - z_k(t)) , \quad (23.20)$$

where we take the convention  $\mathbf{z}_k \neq \mathbf{z}_m$  if  $k \neq m$ . Using above current field, we can change the free particle's action into the form:

$$L = \int d^D x \left[ \frac{m}{2\rho} \mathbf{j}^2 + \alpha(\nabla \cdot \mathbf{j} + \dot{\rho}) \right] , \quad (23.21)$$

where the first term reduces to the original Lagrangian by putting the form of current and after the space integration. The second term insures the current conservation and  $\alpha$  is the Lagrange multiplier field. Since this action has no local gauge invariance, the system is second class and its constraint analysis is done straightforward.

The Euler-Lagrange equation leads to three equations

$$\nabla \cdot \mathbf{j} + \dot{\rho} = 0 , \quad \nabla \alpha = \frac{m}{\rho} \mathbf{j} , \quad \dot{\alpha} = -\frac{m}{2\rho^2} \mathbf{j}^2 . \quad (23.22)$$

By using the second equation of above ones, we can change the Lagrangian into the form

$$L = \int d^D x \left[ -\rho \dot{\alpha} - \frac{\rho}{2m} (\nabla \alpha)^2 \right] . \quad (23.23)$$

This is equal to the Lagrangian for Schrödinger field

$$L = \int d^D x \left[ i\hbar \Psi^* \partial_t \Psi - \Psi^* \hat{H} \Psi \right] , \quad \hat{H} = -\frac{\hbar^2}{2m} \nabla^2 , \quad (23.24)$$

with

$$\Psi \sim \sqrt{\rho(t)} e^{i\alpha(x,t)/\hbar} , \quad (23.25)$$

where we take the approximation that  $\rho$  changes slowly in space, and we have neglected the total derivative. From (23.24) we can find correspondence between Lagrange multiplier field  $\alpha$  and quantum mechanical phase variable.

Now we go back to Lagrangian (23.21). This Lagrangian has the symmetry of global translation for multiplier field.

$$\delta\alpha(x, t) = C , \quad (23.26)$$

where  $C$  is constant. The Noether current for this translation is the current appearing in action as dynamical variable. Let us extend this translation to the local one. The local translation,

$$\delta\alpha(x, t) = \theta(x, t) , \quad (23.27)$$

does not change the action only when we extend the derivative for  $\alpha$  field as

$$\partial_\mu\alpha(x, t) \rightarrow D_\mu\alpha(x, t) \equiv \partial_\mu\alpha(x, t) + A_\mu(x, t) , \quad (23.28)$$

where the transformation law for  $A_\mu$  is

$$\delta A_\mu(x, t) = -\partial_\mu\theta(x, t) . \quad (23.29)$$

We used here the relativistic notation. Notice that this is not the usual covariant derivative, and  $D_\mu\alpha(x, t)$  is not only covariant but also invariant under the local translation. Using the notation

$$j^\mu = (\rho, j^1, j^2, j^3) , \quad (23.30)$$

our local gauge invariant action takes the form:

$$S = \int d^D x dt \left[ \frac{m}{2\rho} \mathbf{j}^2 - j^\mu A_\mu + \alpha \partial_\mu j^\mu \right] . \quad (23.31)$$

So we get the usual gauge coupling.

Solving the constraint (current conservation), we find the solution as the form of currents (23.20), and by putting them into the above action we obtain the gauge coupled action for particles.

$$S = \int dt \sum_{k=1}^N \left[ \frac{m}{2} \dot{\mathbf{z}}_k^2 - \dot{z}_k^i A_i(z_k, t) - A_0(z_k, t) \right] . \quad (23.32)$$

In this way we can introduce  $U(1)$  gauge field by gauge principle completely in classical theory. The local symmetry is not the  $U(1)$  transformation but 1-dimensional translation. The reason why the multiplier field corresponds to the phase is the following. In quantum theory the translation of phase induced the current conservation, and in our case the same job is done by the multiplier field. So they are corresponding each other in the sense of gauge transformation.

### 23.6.1 Field Equations

We consider the free field equation given by the action,

$$S = \int d^D x dt \left[ -\rho \dot{\alpha} - \frac{\rho}{2m} (\nabla \alpha)^2 \right]. \quad (23.33)$$

The Euler-Lagrange equations give

$$\begin{aligned} \dot{\alpha} + \frac{1}{2m} (\nabla \alpha)^2 &= 0, \\ \dot{\rho} + \frac{1}{m} \nabla \cdot (\rho \nabla \alpha) &= 0. \end{aligned} \quad (23.34)$$

The solution of the first non-linear equation has the form:

$$\alpha(x, t) = 2mC (|\vec{x} - \vec{x}_0| - Ct) + D, \quad (23.35)$$

where  $\vec{x}_0$ ,  $C$ ,  $D$  are the integration constant. By using the notation  $r = |\vec{x} - \vec{x}_0|$ , the solution of the second equation in (23.34) is given as,

$$\rho(x, t) = \int d\omega \frac{A(\omega)}{r^2} e^{i\omega(t-r/2C)}. \quad (r \neq 0) \quad (23.36)$$

The green function, which is described by

$$\rho \dot{G} + \frac{1}{m} \nabla \cdot (\rho_G \nabla \alpha) = \delta^3(\vec{x} - \vec{x}_0) \delta(t), \quad (23.37)$$

is given as

$$\rho_G(\vec{x}, \vec{x}_0, t) = \frac{1}{8\pi C \sqrt{2mr^2}} \delta\left(t - \frac{r}{2C}\right). \quad (23.38)$$

This solution says that particle density moves at the constant speed  $2C$  which is determined by the initial condition. This is essentially classical picture, and much different from the Schrödinger equation, where the equation is diffusion type and propagation belongs to that one. If we work with Schrödinger equation, we should add  $-\hbar^2(\nabla\rho)^2/8m\rho$  to the action (23.33). Then the first equation of (23.34) changes to include  $\rho$  dependent term to be highly non-linear equation, even though the Schrödinger equation for the function (23.25) is simple linear one.

### 23.6.2 SSB and Higgs Mechanism

Let us include the gauge field into the action as above,

$$S = \int d^D x dt \left[ -\rho \dot{\alpha} - \frac{\rho}{2m} (\nabla \alpha - \mathbf{A}(x, t))^2 + \rho A^0 \right] , \quad (23.39)$$

where we write  $\mathbf{A} = A^k = -A_k$  in usual relativistic manner. The Noether current corresponds to the translation invariance is

$$\begin{aligned} j^0 &= \rho , \\ \mathbf{j} &= \frac{\rho}{m} (\nabla \alpha - \mathbf{A}) . \end{aligned} \quad (23.40)$$

Therefore the Noether charge  $Q$  is given by

$$Q = \int d^3 x \rho(x) , \quad (23.41)$$

and induce the infinitesimal translation of  $\alpha$ .

Let us consider the quantization of this field theory. The canonical momentum conjugate to density  $\rho$  is the "classical-phase"  $\alpha$  and we get

$$[\rho(x), \alpha(y)] = i\hbar \delta^3(x - y) . \quad (23.42)$$

This relation induce the relation

$$\langle 0|[Q, \alpha(x)]|0\rangle = i\hbar = \text{const} . \quad (23.43)$$

This means the SSB and Nambu Goldstone boson is the  $\alpha$  field.

If we add the kinetic term of gauge field in (23.33), we can perform the gauge transformation for gauge field freely. The gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha , \quad (23.44)$$

"gauge out" the  $\alpha$  field, and the action (23.33) takes the form:

$$S = \int d^D x dt \left[ -\frac{\rho}{2m} \mathbf{A}^2(x, t) - \rho A^0 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \right] . \quad (23.45)$$

$\rho$  is no more dynamical variable. The equation for gauge field shows that it gets the mass

$$M = \sqrt{\frac{\rho q^2}{m}} , \quad (23.46)$$

where we add the electric charge  $q$  to the gauge coupling. This is the plasma frequency.

Even if we do not gauge out the  $\alpha$  field, the electric current has the form:

$$\mathbf{j}_q = -\frac{\bar{\rho}q^2}{m}(\mathbf{A} - \nabla\alpha), \quad (23.47)$$

and taking rotation, we obtain

$$\nabla \times \mathbf{j}_q = -\frac{\bar{\rho}q^2}{m}\mathbf{B}. \quad (23.48)$$

This is the same as the London equation for super-conductors. Here we gave the condition by hand

$$\rho = \bar{\rho} + \delta\rho \sim \bar{\rho} = \text{const}. \quad (23.49)$$

So we have the anti-magnetism just like super-conductivity. Note that this analogy is not complete, our classical field theory should be compared to the charged boson models rather than super-conductivity theory, since for fermions the *ansatz* (23.25) cannot be applied.

**Exercise 23.1:** Within the scalar QED with Higgs phenomena consider the static case  $\partial_t\phi = \partial_t A^i = 0$  and  $A^0 = 0$ .

- (a) Find the equation of motion for  $A^i$  and expression for the current with SSB (the London equation), from which follows the Meissner effect;
- (b) The resistivity for the system  $\rho$  is defined by  $\vec{E} = \rho\vec{J}$ . Show that, in this case of spontaneous symmetry breaking,  $\rho = 0$ , and we have superconductivity.

## Chapter 24

# Higgs Mechanism in SM

In SM gauge invariance (and therefore renormalizability) does not allow mass terms in the Lagrangian for the gauge bosons or for chiral fermions. Massless gauge bosons are not acceptable for the weak interactions, which are known to be short-ranged. Hence, the gauge invariance must be broken spontaneously, which preserves the renormalizability. The idea is that the lowest energy (vacuum) state does not respect the gauge symmetry and induces effective masses for particles propagating through it.

### 24.1 Gauge Bosons Masses

Let us introduce the complex vector

$$v = \langle 0|\phi|0\rangle = \text{const} , \tag{24.1}$$

which has components that are *the Vacuum Expectation Values* (VEV) of the various complex scalar fields. As it was discussed in previous sections  $v$  is determined by rewriting the Higgs potential as a function of  $v$ ,

$$V(\phi) \rightarrow V(v) , \tag{24.2}$$

and choosing  $v$  such that  $V$  is minimized. That is, we interpret  $v$  as the lowest energy solution of the classical equation of motion. The quantum theory is obtained by considering fluctuations around this classical minimum,

$$\phi = v + \phi' . \tag{24.3}$$

Note that it suffices to consider constant  $v$  because any space or time dependence of  $v$  would increase the energy of the solution. Also, one can take

$$\langle 0|A_\mu|0\rangle = 0 , \quad (24.4)$$

because any non-zero vacuum value for a higher-spin field would violate Lorentz invariance. However, these extensions are involved in higher energy classical solutions (topological defects), such as monopoles, strings, domain walls, and textures.

The single complex Higgs doublet in the SM can be rewritten in a Hermitian basis as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 - i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 - i\phi_4) \end{pmatrix} , \quad (24.5)$$

where

$$\phi_i = \phi_i^\dagger \quad (24.6)$$

represent four Hermitian fields. In this new basis the Higgs potential becomes

$$V(\phi) = \frac{1}{2}\mu^2 \left( \sum_{i=1}^4 \phi_i^2 \right) + \frac{1}{4}\lambda \left( \sum_{i=1}^4 \phi_i^2 \right)^2 , \quad (24.7)$$

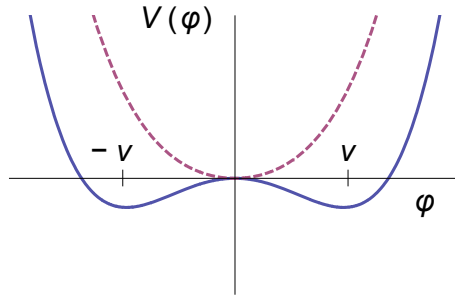
which is clearly  $O(4)$  invariant. Without loss of generality we can choose the axis in this four-dimensional space so that

$$\begin{aligned} \langle 0|\phi_i|0\rangle &= 0 , & (i = 1, 2, 4) \\ \langle 0|\phi_3|0\rangle &= v . \end{aligned} \quad (24.8)$$

Thus,

$$V(\phi) \rightarrow V(v) = \frac{1}{2}\mu^2 v^2 + \frac{1}{4}\lambda v^4 , \quad (24.9)$$

which must be minimized with respect to  $v$ . Two important cases are illustrated in Figure below.





- For  $\mu^2 > 0$  (dashed line) the minimum occurs at  $v = 0$ . That is, the vacuum is empty space and  $SU(2) \times U(1)$  is unbroken at the minimum;
- For  $\mu^2 < 0$  (solid line) the  $v = 0$  symmetric point is unstable, and the minimum occurs at some nonzero value of  $v$  which breaks the  $SU(2) \times U(1)$  symmetry.

The extremum point is found by requiring

$$V'(v) = v(\mu^2 + \lambda v^2) = 0 , \quad (24.10)$$

which has the solution

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \quad (24.11)$$

at the minimum. The solution for  $-v$  can also be transformed into this standard form by an appropriate  $O(4)$  transformation. The dividing point  $\mu^2 = 0$  cannot be treated classically. It is necessary to consider the one loop corrections to the potential, in which case it is found that the symmetry is again spontaneously broken.

We are interested in the case  $\mu^2 < 0$ , for which the Higgs doublet is replaced by its classical value (in first approximation),

$$\phi \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \equiv v . \quad (24.12)$$

The isotopical spin generators  $L^1$  and  $L^2$ , and the generator  $L^3 - Y$  are spontaneously broken, e.g.

$$L^1 v \neq 0 . \quad (24.13)$$

On the other hand, the vacuum carries no electric charge,

$$Qv = (L^3 + Y)v = 0 , \quad (24.14)$$

so the  $U(1)_{em}$  of electromagnetism is not broken. Thus, the EW  $SU(2) \times U(1)$  group is spontaneously broken to the  $U(1)_{em}$  subgroup,

$$SU(2) \times U(1)_Y \rightarrow U(1)_{em} . \quad (24.15)$$

To quantize around the classical vacuum, write

$$\phi = v + \phi' , \quad (24.16)$$

where  $\phi'$  are quantum fields with zero vacuum expectation value. To display the physical particle content it is useful to rewrite the four Hermitian components of  $\phi'$  in terms of a new set of variables using the Kibble transformation:

$$\phi = \frac{1}{\sqrt{2}} e^{i \sum \xi^i L^i} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (24.17)$$

$H$  is a Hermitian field which will turn out to be the physical Higgs scalar.

If we had been dealing with a global SSB the three Hermitian fields  $\xi^i$  would be the massless pseudo scalar Nambu-Goldstone bosons that are necessarily associated with broken symmetry generators. However, in a gauge theory they disappear from the physical spectrum. To see this it is useful to go to the unitary gauge

$$\phi \rightarrow \phi' = e^{-i \sum \xi^i L^i} \phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \quad (24.18)$$

in which the Goldstone bosons disappear. In this gauge, the scalar covariant kinetic energy term takes the simple form

$$\begin{aligned} (D_\mu \phi)^\dagger D^\mu \phi &= \frac{1}{2} (0 \ v) \left[ \frac{g_2}{2} \tau^i W_\mu^i + \frac{g_1}{2} B_\mu \right]^2 \begin{pmatrix} 0 \\ v \end{pmatrix} + H \text{ terms}, \\ &\Rightarrow M_W^2 W^{+\mu} W_\mu^- + \frac{M_Z^2}{2} Z^\mu Z_\mu + H \text{ terms}, \end{aligned} \quad (24.19)$$

where the kinetic energy and gauge interaction terms of the physical  $H$  particle have been omitted. Thus, SSB generates mass terms for the  $W$  and  $Z$  gauge bosons

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}} (W^1 \mp iW^2), \\ Z &= -\sin \theta_W B + \cos \theta_W W^3. \end{aligned} \quad (24.20)$$

The photon field

$$A = \cos \theta_W B + \sin \theta_W W^3 \quad (24.21)$$

remains massless. The masses are

$$M_W = \frac{g_2 v}{2} \quad (24.22)$$

and

$$M_Z = \sqrt{g_1^2 + g_2^2} \frac{v}{2} = \frac{M_W}{\cos \theta_W}, \quad (24.23)$$

where the weak angle is defined by

$$\tan \theta_W \equiv \frac{g_1}{g_2} \quad \Rightarrow \quad \sin^2 \theta_W = 1 - \frac{M_W^2}{M_Z^2} . \quad (24.24)$$

One can think of the generation of masses as due to the fact that the  $W$  and  $Z$  interact constantly with the condensate of scalar fields and therefore acquire masses, in analogy with a photon propagating through a plasma. The Goldstone boson has disappeared from the theory but has reemerged as the longitudinal degree of freedom of a massive vector particle.

It will be seen below that

$$\frac{G_F}{\sqrt{2}} \sim \frac{g_2^2}{8M_W^2} , \quad (24.25)$$

where

$$G_F = 1.16637(5) \times 10^{-5} \text{ GeV}^{-2} \quad (24.26)$$

is the Fermi constant determined by the muon lifetime. The weak scale  $v$  is therefore

$$v = \frac{2M_W}{g} \simeq (\sqrt{2}G_F)^{-1/2} \simeq 246 \text{ GeV} . \quad (24.27)$$

Similarly,

$$g_2 = \frac{e}{\sin \theta_W} , \quad (24.28)$$

where  $e$  is the electric charge of the positron. Hence, to lowest order

$$M_W = M_Z \cos \theta_W \sim \frac{(\pi\alpha/\sqrt{2}G_F)^{1/2}}{\sin \theta_W} , \quad (24.29)$$

where

$$\alpha \sim \frac{1}{137.036} \quad (24.30)$$

is the fine structure constant. Using

$$\sin^2 \theta_W \sim 0.23 , \quad (24.31)$$

obtained from NC scattering, one expects

$$M_W \sim 78 \text{ GeV}, \quad M_Z \sim 89 \text{ GeV} . \quad (24.32)$$

These predictions are increased by  $\sim (2 - 3)$  GeV by loop corrections.

The  $W$  and  $Z$  were discovered at CERN in 1983. Subsequent measurements of their masses and other properties have been in excellent agreement with the SM expectations (including the higher-order corrections). The current values are

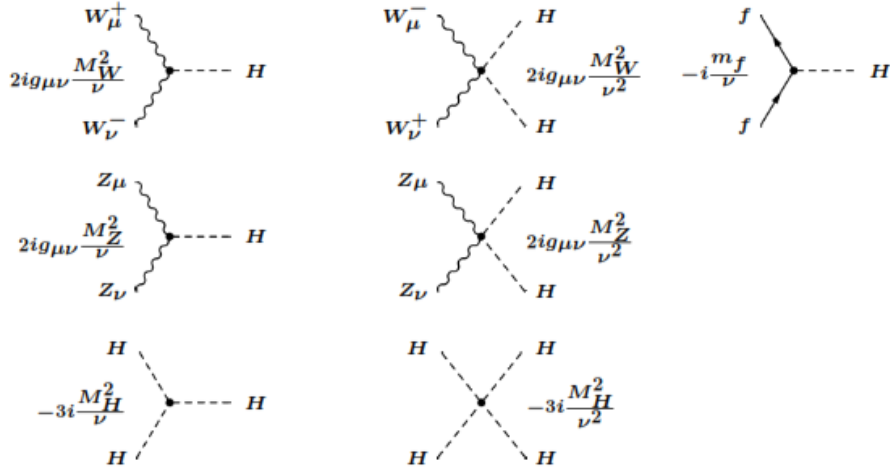
$$M_W = 80.398 \pm 0.025 \text{ GeV} , \quad M_Z = 91.1876 \pm 0.0021 \text{ GeV} . \quad (24.33)$$

## 24.2 The Higgs Sector in Lagrangian

The Higgs is a scalar field in the  $(\mathbf{2}, -1/2)$  representation of  $SU(2) \otimes U(1)$ . We know that the Higgs Lagrangian will have the kinetic term and some potential:

$$\begin{aligned} \mathcal{L}_\phi = & (D^\mu \phi)^\dagger D_\mu \phi - V(\phi) = M_W^2 W^{\mu+} W_\mu^- \left(1 + \frac{H}{v}\right)^2 + \\ & + \frac{1}{2} M_Z^2 Z^\mu Z_\mu \left(1 + \frac{H}{v}\right)^2 + \frac{1}{2} (\partial_\mu H)^2 - V(\phi). \end{aligned} \quad (24.34)$$

This includes the  $W$  and  $Z$  mass terms and also the  $ZZH^2$ ,  $W^+W^-H^2$  and the induced  $ZZH$  and  $W^+W^-H$  interactions (as shown in the figure below). The last two terms are the canonical Higgs kinetic energy term and the potential.



After symmetry breaking the Higgs potential in unitary gauge becomes

$$V(\phi) = -\frac{\mu^4}{4\lambda} - \mu^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4. \quad (24.35)$$

The first term in the Higgs potential  $V$  is a constant,

$$\langle 0|V(v)|0\rangle = -\frac{\mu^4}{4\lambda}. \quad (24.36)$$

It reflects the fact that  $V$  was defined so that

$$V(0) = 0 , \quad (24.37)$$

and therefore

$$V < 0 \quad (24.38)$$

at the minimum. Such a constant term is irrelevant to physics in the absence of gravity, but will be seen later to be one of the most serious problems of the SM when gravity is incorporated because it acts like a cosmological constant much larger (and of opposite sign) than is allowed by observations.

The second term in the Higgs potential (24.35) represents a (tree-level) mass

$$M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda} v , \quad (24.39)$$

for the Higgs boson. The weak scale is given in (24.27), but the quartic Higgs coupling  $\lambda$  is unknown, so  $M_H$  is not predicted. A priori,  $\lambda$  could be anywhere in the range  $0 \leq \lambda < \infty$ , but from the observed Higgs mass value,  $M_H \approx 125$  GeV, from (24.27) and (24.39) one can obtain  $\lambda \approx 0.13$ .

The third and fourth terms in (24.35) represent the induced cubic and quartic interactions of the Higgs scalar, shown in the table and figure above.

### 24.3 Yukawa Interactions

The Yukawa interaction in the unitary gauge becomes

$$\begin{aligned} -\mathcal{L}_{\text{Yukawa}} &\rightarrow \sum_{m,n=1}^N \bar{u}_{mL}^0 \Gamma_{mn}^u \left( \frac{v+H}{\sqrt{2}} \right) u_{mR}^0 + (d, e, \nu) \text{ terms} + h.c. = \\ &= \bar{u}_L^0 (M^u + h^u H) u_R^0 + (d, e, \nu) \text{ terms} + h.c. , \end{aligned} \quad (24.40)$$

where

$$u_L^0 = (u_{1L}^0 u_{2L}^0 \cdots u_{FL}^0)^T \quad (24.41)$$

is an  $N$ -component column vector, with a similar definition for  $u_R^0$ .  $M^u$  is an  $N \times N$  fermion mass matrix

$$M_{mn}^u = \frac{1}{\sqrt{2}} \Gamma_{mn}^u v \quad (24.42)$$

induced by spontaneous symmetry breaking, and

$$h^u = \frac{M^u}{v} = \frac{g_2 M^u}{2M_W} \quad (24.43)$$

is the Yukawa coupling matrix.

In general  $M$  is not diagonal, Hermitian, or symmetric. To identify the physical particle content it is necessary to diagonalize  $M$  by separate unitary transformations  $A_L$  and  $A_R$  on the left- and right-handed fermion fields. In the special case that  $M^u$  is Hermitian one can take

$$A_L = A_R . \quad (24.44)$$

Then,

$$A_L^{u\dagger} M^u A_R^u = M_D^u = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \quad (24.45)$$

is a diagonal matrix with eigenvalues equal to the physical masses of the charge-2/3 quarks.

From (24.45) and its conjugate one has

$$\hat{A}_L^{u\dagger} M^u M^{u\dagger} \hat{A}_L^u = \hat{A}_R^{u\dagger} M^{u\dagger} M^u \hat{A}_R^u = M_D^{u2} . \quad (24.46)$$

But  $MM^\dagger$  and  $M^\dagger M$  are Hermitian, so  $A_{L,R}$  can then be constructed by elementary techniques, up to overall phases that can be chosen to make the mass eigenvalues real and positive, and to remove unobservable phases from the weak CC.

Similarly, one diagonalizes the down quark, charged lepton, and neutrino mass matrices by

$$\begin{aligned} A_L^{d\dagger} M^d A_R^d &= M_D^d , \\ A_L^{e\dagger} M^e A_R^e &= M_D^e , \\ A_L^{\nu\dagger} M^\nu A_R^\nu &= M_D^\nu . \end{aligned} \quad (24.47)$$

In terms of these unitary matrices we can define mass eigenstate fields

$$u_L = A_L^{u\dagger} u_L^0 = (u_L \ c_L \ t_L)^T , \quad (24.48)$$

with analogous definitions for

$$\begin{aligned} u_R &= A_R^{u\dagger} u_R^0 , & d_{L,R} &= A_{L,R}^{d\dagger} d_{L,R}^0 , \\ e_{L,R} &= A_{L,R}^{e\dagger} e_{L,R}^0 , & \nu_{L,R} &= A_{L,R}^{\nu\dagger} \nu_{L,R}^0 . \end{aligned} \quad (24.49)$$

Typical estimates of the quark masses are

$$\begin{aligned} m_u &\sim 1.5 - 3 \text{ MeV} , & m_d &\sim 3 - 7 \text{ MeV} , \\ m_s &\sim 70 - 120 \text{ MeV} , & m_c &\sim 1.5 - 1.8 \text{ GeV} , \\ m_b &\sim 4.7 - 5.0 \text{ GeV} , & m_t &= 170.9 \pm 1.8 \text{ GeV} . \end{aligned} \quad (24.50)$$

These are the current masses: for QCD their effects are identical to bare masses in the QCD Lagrangian. They should not be confused with the constituent masses of order 300 MeV generated by the spontaneous breaking of chiral symmetry in the strong interactions. Including QCD renormalizations, the  $u$ ,  $d$ , and  $s$  masses are running masses evaluated at  $2 \text{ GeV}^2$ , while  $m_{c,b,t}$  are pole masses.

So far we have only allowed for ordinary Dirac mass terms of the form  $\bar{\nu}_{mL}^0 \nu_{nR}^0$  for the neutrinos, which can be generated by the ordinary Higgs mechanism. Another possibility are lepton number violating Majorana masses, which require an extended Higgs sector or higher-dimensional operators. It is not clear yet whether Nature utilizes Dirac masses, Majorana masses, or both. What is known, is that the neutrino mass eigenvalues are tiny compared to the other masses,  $\lesssim \mathcal{O}(0.1) \text{ eV}$ , and most experiments are insensitive to them. In describing such processes, one can ignore  $\Gamma^\nu$ , and the  $\nu_R$  effectively decouple. Since

$$M^\nu \sim 0 \quad (24.51)$$

the three mass eigenstates are effectively degenerate with eigenvalues 0, and the eigenstates are arbitrary. That is, there is nothing to distinguish them except their weak interactions, so we can simply define  $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$  as the weak interaction partners of the  $e$ ,  $\mu$ , and  $\tau$ , which is equivalent to choosing

$$A_L^\nu \equiv A_L^e \quad (24.52)$$

so that

$$\nu_L = A_L^{e\dagger} \nu_L^0 . \quad (24.53)$$

Of course, this is not appropriate for physical processes, such as oscillation experiments, that are sensitive to the masses or mass differences.

In terms of the mass eigenstate fermions,

$$-\mathcal{L}_{\text{Yukawa}} = \sum_i m_i \bar{\psi}_i \psi_i \left( 1 + \frac{g_2}{2M_W} H \right) = \sum_i m_i \bar{\psi}_i \psi_i \left( 1 + \frac{H}{v} \right) . \quad (24.54)$$

The coupling of the physical Higgs boson to the  $i^{\text{th}}$  fermion is  $g_2 m_i / 2M_W$ , which is very small except for the top quark. The coupling is flavor-diagonal in the minimal model: there is just one Yukawa matrix for each type of fermion, so the mass and Yukawa matrices are diagonalized by the same transformations.

In generalizations in which more than one Higgs doublet couples to each type of fermion there will in general be flavor-changing Yukawa interactions involving the physical neutral Higgs fields. There are stringent limits on such couplings; for example, the  $K_L - K_S$  mass difference implies

$$\frac{h}{M_H} < 10^{-6} \text{ GeV}^{-1}, \quad (24.55)$$

where  $h$  is the  $\bar{d}s$  Yukawa coupling.

**Exercise 24.1:** Let  $\phi_i$  be the scalar fields in the vector representation of the  $SU(n)$  group.

- Write down the  $SU(n)$  invariant scalar potential for  $\phi_i$ ;
- Work out the possible pattern for the spontaneous symmetry breaking for  $\phi_i$ . How many Goldstone bosons are there in this case?
- Discuss the possible spontaneous symmetry breaking pattern for the case where there are two such scalar fields  $\phi_i^1$  and  $\phi_i^2$ .

**Exercise 24.2:** Suppose  $\phi_i^j$  are scalar fields in the adjoint representation of an  $SU(n)$  group.

- Write down the scalar potential for  $\phi_i^j$ ;
- Work out the possible pattern for the SSB for  $\phi_i^j$ .

**Exercise 24.3:** According to the Weinberg-Salam model, the Higgs boson  $H$  couples to every elementary fermion  $\psi$  in the form:  $em_\psi/m_W H\bar{\psi}\psi$ , where  $e$  is the charge of the electron, and  $m_W$  is the mass of the  $W$  boson. Assuming that the Higgs boson decays primarily to the known quarks and leptons, calculate its lifetime in terms of the Higgs mass.

**Exercise 24.4:** Find the  $HZZ$  coupling.

**Exercise 24.5:** Draw the lowest-order Feynman diagrams for the processes  $e^+e^- \rightarrow HZ$  and  $e^+e^- \rightarrow H\nu_e\bar{\nu}_e$ , which are the main Higgs production mechanism at a future high-energy linear collider.

**Exercise 24.6:** Calculate the partial decay width of a Higgs boson into a fermion-antifermion pair.



## Part IX

# Lecture – Summary of the SM and Predictions



# Chapter 25

## The SM Lagrangian

As it was already mentioned in previous lectures the SM is based on the  $SU(3) \times SU(2) \times U(1)$  Lagrangian, which can be written in the form:

$$\mathcal{L}_{SU(3) \times SU(2) \times U(1)} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\phi} + \mathcal{L}_f + \mathcal{L}_{\text{Yukawa}} . \quad (25.1)$$

The strong interaction part (QCD) is an  $SU(3)$  gauge theory, which will be considered in the next lecture. Here we will descuse in details the four constituents of (25.1) for EW, i.e.  $SU(2) \times U(1)$  theory.

### 25.1 The Gauge Part

The gauge part of the the SM Lagrangian (25.1) can be written as the sum of the kinetic terms for the gauge bosons of the QCD and EW model,

$$\mathcal{L}_{\text{gauge}} = \mathcal{L}_{\text{gauge}}^{\text{QCD}} + \mathcal{L}_{\text{gauge}}^{\text{EW}} . \quad (25.2)$$

QCD will be considered in the next lecture. The EW gauge part (without gluons),

$$\mathcal{L}_{\text{gauge}}^{\text{EW}} = -\frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} , \quad (a = 1, 2, 3) \quad (25.3)$$

contains the  $SU(2)$  and  $U(1)$  strength tensors:

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu , \\ W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_2 \epsilon_{abc} W_\mu^b W_\nu^c , \end{aligned} \quad (25.4)$$

where  $W_\mu^a$  and  $B_\mu$  are respectively the corresponding gauge fields,  $g_2$  is the  $SU(2)$  gauge coupling and  $\epsilon_{abc}$  is the totally antisymmetric symbol. The coupling for the  $U(1)$  gauge field  $B_\mu$  will be denoted below by  $g_1$ . The  $SU(2)$  fields,  $W_\mu^a$ , have three and four-point self-interactions.  $B_\mu$  has no self-interactions and is a  $U(1)$  field associated with the weak hypercharge

$$Y = Q - T_L^3, \quad (25.5)$$

where  $Q$  and  $T_L^3$  are respectively the electric charge operator and the third component of the weak  $SU(2)$  isotopic spin. The  $B^\mu$  and  $W_3^\mu$  fields will eventually mix to form the photon and  $Z$  boson.

## 25.2 The Scalar Part

The scalar part of the Lagrangian (25.1) is:

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger D_\mu \phi - V(\phi), \quad (25.6)$$

where  $\phi$  is a complex Higgs scalar, which is a color singlet and an  $SU(2)_L$  doublet, given by

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (25.7)$$

i.e. we have four real scalar fields.

In (25.6)  $V(\phi)$  is the Higgs potential. The combination of  $SU(2) \times U(1)$  invariance and renormalizability restricts  $V$  to the form

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (25.8)$$

As we already noted in the previous lecture, compared to the Lagrangian for a free complex scalar field, this potential has the wrong sign for the quadratic term. So  $\mu$  is not the mass and we cannot interpret the excitations of the field  $\phi$  as propagating degrees of freedom. But it is precisely this wrong sign that is required for the SSB to occur. For the quartic self-interaction between the scalar fields vacuum stability requires  $\lambda > 0$ .

The minimum of the potential (25.8) occurs for

$$\phi^\dagger \phi = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \quad (25.9)$$

The  $SU(2)$  symmetry is broken when the vacuum field configuration chooses a particular direction in the  $\phi^1, \phi^2$  space. The choice of the representation of the Higgs field decides pattern of symmetry breaking. For the case under consideration,  $SU(2)_L \times U(1)_Y$ , the unbroken symmetry should correspond to the  $U(1)_{em}$  invariance since the photon is massless. Glashow's partial symmetry breaking formula (25.5) helps us decide which of the four scalar fields can acquire a nonzero VEV. The charge operator  $Q$  should annihilate the vacuum and hence only the electrically neutral, real scalar field can have a nonzero VEV. The required symmetry breaking pattern is guaranteed (with the choice  $Y_\phi = 1$ ) by

$$\langle 0|\phi|0\rangle = \langle \phi \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (25.10)$$

As follows from (25.9),  $v = \mu^2/\sqrt{\lambda}$ . Since  $\phi$  is a  $SU(2)_L$  doublet, this choice for the VEV means that the vacuum configuration breaks the symmetry and chooses a particular minimum from amongst the continuum of minima. Since the electromagnetic charge still annihilates the vacuum, the symmetry breaking pattern is

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}. \quad (25.11)$$

One can rewrite  $\phi(x)$  in terms of some fields  $\theta_a(x)$  and  $h(x)$  (all of which have vacuum expectation value to be 0),

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_2 + i\theta_1 \\ v + h - i\theta_3 \end{pmatrix}. \quad (25.12)$$

If  $\theta_a(x)$  and  $h(x)$  are small then we get

$$\phi(x) = \frac{1}{\sqrt{2}} e^{i\theta_a \tau^a / v} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (25.13)$$

This is then an expansion of the field  $\phi$  in terms of the fluctuations around the minimum. One recognizes the exponential factor outside as that for a gauge transformation for a  $SU(2)_L$  doublet. We see that by doing a gauge transformation

$$\phi' = -e^{i\theta_a \tau^a / v} \phi \quad (25.14)$$

we get,

$$\phi'(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (25.15)$$

This gauge is called the Unitary gauge, (25.10) also means that the VEV is zero for  $h(x)$ . The three scalar degrees of freedom  $\theta_a$  in fact have disappeared from the spectrum in this gauge. Indeed these three fields correspond to three Goldstone bosons corresponding to the three generators of the symmetry group that are broken spontaneously.

### 25.2.1 Masses of Gauge Bosons

Let us now evaluate the kinetic term of the scalar Lagrangian (25.6) in the unitary gauge using  $\phi'$  from (25.15). We use

$$D_\mu \phi = \partial_\mu \phi - i \frac{g_1}{2} B_\mu \phi - i g_2 W_\mu^a \frac{\tau^a}{2} \phi. \quad (25.16)$$

This covariant derivative gives rise to terms quadratic in the gauge boson fields which are given as below:

$$\begin{aligned} & \left| \left( \frac{g_1}{2} B_\mu + g_2 \frac{\tau^a}{2} W_\mu^a \right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 = \\ &= \frac{g_2^2 v^2}{8} (W_\mu^a W^{a\mu}) + \frac{g_1^2 v^2}{8} B_\mu B^\mu - \frac{g_1 g_2 v^2}{4} W_\mu^3 B^\mu = \\ &= \frac{g_2^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{v^2}{8} (g_1 B_\mu - g_2 W_\mu^3)^2 = \\ &= \frac{g_2^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{(g_1^2 + g_2^2) v^2}{8} Z_\mu Z^\mu. \end{aligned} \quad (25.17)$$

So three of the four gauge bosons become massive: the  $W_\mu^\pm$  and one linear combination of  $B_\mu$  and  $W_\mu^3$ , which we call  $Z_\mu$ , and the orthogonal linear combination  $A_\mu$  remains massless. This also tells us

$$\begin{aligned} M_W^2 &= \frac{g_2^2 v^2}{4}, \\ M_Z^2 &= \frac{(g_1^2 + g_2^2) v^2}{4} = \frac{M_W^2}{\cos^2 \theta_W}, \end{aligned} \quad (25.18)$$

where the weak angle  $\theta_W$  is defined by:

$$\tan \theta_W \equiv \frac{g_1}{g_2} \quad \Rightarrow \quad \sin^2 \theta_W = 1 - \frac{M_W^2}{M_Z^2}. \quad (25.19)$$

Another fact worth noticing is that the value of the VEV,  $v$ , gets determined in terms of measured value of  $G_\mu$ . Using the expression for  $M_W$  in (25.18) and that for

$$G_\mu \approx 1.17 \times 10^{-5} \text{ GeV}^{-2}, \quad (25.20)$$

we get

$$v = \left( \frac{1}{\sqrt{2}G_\mu} \right)^{1/2} \simeq 246 \text{ GeV}. \quad (25.21)$$

Using the expression

$$g_2 = \frac{e}{\sin \theta_W} \quad (25.22)$$

in terms of the positron charge  $e$  and  $\sin \theta_W$  and (25.18), one can then see that,

$$\begin{aligned} M_W &= \sqrt{\frac{\pi}{\sqrt{2}G_\mu} \frac{\alpha_{\text{em}}}{\sin^2 \theta_W}} = \frac{37.3}{\sin \theta_W} \text{ GeV}; \\ M_Z &= \frac{37.3}{\sin \theta_W \cos \theta_W} \text{ GeV}. \end{aligned} \quad (25.23)$$

Now everything in the EW model is predicted in terms of the two known constants  $\alpha_{\text{em}}$ ,  $G_\mu$  and one free parameter  $\sin^2 \theta_W$ .

We further notice from (25.18) that the ratio

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} \quad (25.24)$$

is predicted to be unity in the EW model.

We can conclude that one should expect the neutrino induced scattering processes via NC interactions to happen at rates similar to those via CC interactions. This conclusion is of course independent of the actual values of  $M_W$  and  $M_Z$  with the proviso that the energies are much smaller compared to these masses. Thus the EW model not only predicted the existence of a weak neutral gauge boson and weak neutral current processes mediated by it, but it also predicted their strength to be  $\mathcal{O}(G_\mu)$ .

After working out the remaining terms also in terms of the field  $\phi'$  in the unitary gauge we get,

$$\begin{aligned} \mathcal{L}_\phi^U &= \left[ M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu \right] \left( 1 + \frac{h}{v} \right)^2 + \\ &+ \frac{1}{2} (\partial_\mu h)^2 + \mu^2 h^2 - \lambda v h^3 - \lambda/4 h^4 = \mathcal{L}_{Vh} + \mathcal{L}_h. \end{aligned} \quad (25.25)$$

The first two terms are the mass terms for the  $W$  and  $Z$ , as well as the term describing the interaction between a pair of gauge bosons and the  $h$ . The form of this term makes it very clear that the strength of the  $VVh$  coupling is simply proportional to the mass of the corresponding gauge boson. This proportionality between the mass and the coupling is the most critical prediction of the SSB.

The remaining terms describe now a real, scalar field which is a propagating degree of freedom with mass  $M_h = \sqrt{2\mu^2}$ . Since  $v = \sqrt{\mu^2/\lambda}$ , the mass of the Higgs boson is given in terms of self-coupling  $\lambda$ . This being an arbitrary parameter of the Higgs potential, not fixed by any condition,  $M_h$  too is a free parameter of the SM, with no prediction for it.

In the unitary gauge now the propagating degrees of freedom are the three massive gauge bosons  $W^\pm$ ,  $Z$ , one massless gauge boson  $\gamma$  and one propagating massive scalar. A massless vector boson has two degrees of freedom corresponding to the two degrees of polarization it can have whereas a massive gauge boson has three degrees of freedom as it can also have longitudinal polarization. Out of the four scalar degrees of freedom only one,  $h$ , is left in the particle spectrum and the other three provide the remaining degrees of freedom corresponding to the longitudinal polarization necessary for the three gauge bosons to be massive.

The total number of bosonic degrees of freedom before SSB are twelve: eight corresponding to four massless gauge boson fields  $W_\mu^a$ ,  $B_\mu$  and the four scalars in  $\phi$ .

After the SSB one has again twelve bosonic degrees of freedom: nine corresponding to the three massive gauge bosons  $W^\pm$ ,  $Z$ , two corresponding to the massless photon  $\gamma$  and one corresponding to the massive neutral scalar  $h$ . In the unitary gauge the particle spectrum contains only the physical fields and the Goldstone boson fields  $\theta_a$  ( $a = 1, 2, 3$ ) of (25.12), are absent from the spectrum. The same is depicted somewhat pictorially below:

$\mathcal{L}_{\text{gauge}}^{\text{massless}}$	+	$\mathcal{L}_\phi$		$\mathcal{L}_{\text{gauge}}^{\text{massive}}$	+	$\mathcal{L}_h$
4 massless gauge bosons		4 scalar fields	SSB → Unitary gauge	3 massive, 1 massless gauge bosons		1 physical scalar
8 d.o.f.		4 d.o.f.		11 d.o.f		1 d.o.f.



## 25.3 The Fermion Part

The fermion term in the SM Lagrangian (25.1),

$$\mathcal{L}_f = \mathcal{L}_{\text{kinetic}}^{\text{Leptons}} + \mathcal{L}_{\text{kinetic}}^{\text{Quarks}} , \quad (25.26)$$

is constructed by the kinetic terms of the left(right) chiral projections of leptons and quark,

$$\psi_{L(R)} \equiv \frac{1}{2}(1 \mp \gamma_5)\psi . \quad (25.27)$$

The left-handed quarks and leptons transform as  $SU(2)$  doublets, while the right-handed fields are singlets.

### 25.3.1 The Lepton Sector

A Lepton is a spin-1/2 particle that does not interact with the  $SU(3)$  color group (the strong force). In SM there are six *Flavors* of leptons arranged into three *Families*, or *Generations*. The first generation consists of the electron ( $e$ ) and the electron neutrino ( $\nu_e$ ), the second generation the muon ( $\mu$ ) and the muon neutrino ( $\nu_\mu$ ), and the third the tau ( $\tau$ ) and tau neutrino ( $\nu_\tau$ ). Each family behaves exactly the same way, so we will only discuss one generation in this section ( $e$  and  $\nu_e$ ). To incorporate the physics of the other families, merely change the  $e$  to either a  $\mu$ , or  $\tau$ , and the  $\nu_e$  to a  $\nu_\mu$ , or  $\nu_\tau$ .

The neutrinos don't really interact with anything on their own (which is why they are incredibly difficult to detect). For this reason, neutrinos don't have their own place in a representation of  $SU(3) \times SU(2) \times U(1)$ . Electrons on the other hand, do interact with other things on their own, and we therefore see them in the (1,1) representation. However, the neutrino does interact with other things as part of an  $SU(2)$  doublet with the electron,

$$l = \begin{pmatrix} \nu_e \\ e \end{pmatrix} . \quad (25.28)$$

This is why it is arranged with the electron under the  $(2, -1/2)$  representation of  $SU(2) \times U(1)$ .

To make this point clear let us start with two fields,  $\bar{e}$  (positron) and  $l$ , where  $\bar{e}$  is a single left-handed Weyl field, and  $l$  is defined in (25.28). As we have said,  $l$  is in the  $(2, -1/2)$  representation,  $\bar{e}$  is in the (1,1) representation,

and  $\nu_e$  has no representation of its own. We can write down the covariant derivative for each field,

$$(D_\mu l)_i = \partial_\mu l_i - ig_2 W_\mu^a (T^a)_{ij} l_j - ig_1 B_\mu Y_l l_i, \quad (25.29)$$

$$D_\mu \bar{e} = \partial_\mu \bar{e} - ig_1 B_\mu Y_{\bar{e}} \bar{e}. \quad (25.30)$$

The field  $\bar{e}$  has no  $SU(2)$  term in its covariant derivative because the 1 representation of  $SU(2)$  is the trivial representation – this means it doesn't carry  $SU(2)$  charge. Also, we know for the hyper-charge that

$$Y_l = -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (25.31)$$

and

$$Y_{\bar{e}} = (1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (25.32)$$

Following the Lagrangian for the spin-1/2 fields, we can write out the kinetic term for both (massless) fields:

$$\mathcal{L}_{\text{kinetic}}^{\text{Leptons}} = i l^\dagger \bar{\sigma}^\mu (D_\mu l)_i + i \bar{e}^\dagger \bar{\sigma}^\mu D_\mu \bar{e}. \quad (25.33)$$

Now we want a kinetic term for the neutrino. It is believed that neutrinos are described by Majorana fields, so we begin with the field

$$\mathcal{N}' = \begin{pmatrix} \nu_e \\ \nu_e^\dagger \end{pmatrix}. \quad (25.34)$$

Now, we employ a trick. The kinetic term for Majorana fields has only one term (because Majorana fields have only one Weyl spinor), whereas the Dirac field sums over both Weyl spinors composing it. So, instead of working with the Majorana field  $\mathcal{N}'$ , we can instead work with the Dirac field

$$\mathcal{N} = \begin{pmatrix} \nu_e \\ 0 \end{pmatrix}. \quad (25.35)$$

So, the Dirac kinetic term  $i \bar{\mathcal{N}} \gamma^\mu \partial_\mu \mathcal{N}$  will clearly result in the correct kinetic term from (25.33), or  $i \nu^\dagger \bar{\sigma}^\mu \partial_\mu \nu$ .

Now, continuing with the symmetry breaking, we want to write the covariant derivative (25.29) and (25.30) in terms of our low energy gauge fields.

The gauge fields corresponding to Cartan generators ( $A_\mu$  and  $Z_\mu$ ) act as force carrying particles, but do not change the charge of the particles they interact with. On the other hand, the non-Cartan generators' gauge fields ( $W_\mu^\pm$ ) are force carrying particles which do change the charge of the particle they interact with. Therefore, to make calculations simpler, we will break the covariant derivative up into the non-Cartan part and the Cartan part.

The non-Cartan part of the covariant derivative (25.29) is

$$\begin{aligned} g_2(W_\mu^1 T^1 + W_\mu^2 T^2) &= \frac{1}{2} g_2 \left[ W_\mu^1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + W_\mu^2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right] = \\ &= \frac{1}{2} g_2 \begin{pmatrix} 0 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & 0 \end{pmatrix} = \quad (25.36) \\ &= \frac{g_2}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}, \end{aligned}$$

and the Cartan part is

$$\begin{aligned} g_2 W_\mu^3 T^3 + g_1 B_\mu Y &= \frac{e}{s_w} (s_w A_\mu + c_w Z_\mu) T^3 + \frac{e}{c_w} (c_w A_\mu - s_w Z_\mu) Y = \\ &= e (A_\mu + \cot \theta_w Z_\mu) T^3 + e (A_\mu - \tan \theta_w Z_\mu) Y = \quad (25.37) \\ &= e (T^3 + Y) A_\mu + e (\cot \theta_w T^3 - \tan \theta_w Y) Z_\mu, \end{aligned}$$

where we used the notation

$$s_w = \sin \theta_w, \quad c_w = \cos \theta_w. \quad (25.38)$$

We have noted before that  $A_\mu$  is the photon, or the electromagnetic field, and  $e$  is the electromagnetic charge. Therefore, the linear combination  $T^3 + Y$  must be the generator of electric charge. Notice that the electromagnetic generator is in a linear combination of the two Cartan generators of  $SU(2) \times U(1)$ .

We know that  $T^3 = \sigma^3/2$ , and  $Y_l$  and  $Y_e$  are defined in equations (25.31) and (25.32), so we can write

$$\begin{aligned} T^3 l &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \nu_e \\ -e \end{pmatrix}, \\ Y_l &= -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} \nu_e \\ e \end{pmatrix}. \quad (25.39) \end{aligned}$$

We know that  $\bar{e}$  carries no  $T^3$  charge, so its  $T^3$  eigenvalue is 0, while  $Y_{\bar{e}}$  is +1. So, summarizing all of this,

$$\begin{aligned} T^3 \nu_e &= +\frac{1}{2} \nu_e, & T^3 e &= -\frac{1}{2} e, & T^3 \bar{e} &= 0, \\ Y \nu_e &= -\frac{1}{2} \nu_e, & Y e &= -\frac{1}{2} e, & Y \bar{e} &= +\bar{e}. \end{aligned} \quad (25.40)$$

Then defining the generator of electric charge to be

$$Q \equiv T^3 + Y, \quad (25.41)$$

we have

$$Q \nu_e = 0, \quad Q e = -e, \quad Q \bar{e} = +\bar{e}. \quad (25.42)$$

So the neutrino  $\nu_e$  has no electric charge, the electron  $e$  has negative electric charge, and the antielectron, or positron, has plus one electric charge – all exactly what we would expect.

The primary idea is that electrons/positrons and neutrinos all interact with the  $SU(2) \times U(1)$  gauge particles, the  $W_\mu^\pm$ ,  $Z_\mu$ , and  $A_\mu$ . The  $Z_\mu$  and  $A_\mu$  (the Cartan gauge particles) interact but do not affect the charge. On the other hand, the  $W_\mu^\pm$  act as  $SU(2)$  raising and lowering operators. The  $SU(2)$  doublet state acted on by these raising and lowering operators is the doublet in equation (25.28). The  $W_\mu^+$  interacts with a left-handed electron, raising its electric charge from minus one to zero, turning it into a neutrino. However  $W_\mu^+$  does not interact with left-handed neutrinos. On the other hand,  $W_\mu^-$  will lower the electric charge of a neutrino, making it an electron. But  $W_\mu^-$  will not interact with an electron. This does not mean that no vertex in the Feynman diagrams will include a  $W_\mu^-$  and an electron field, but rather that if you collide an electron and a  $W_\mu^-$ , there will be no interaction.

### 25.3.2 The Quark Sector

A Quark is a spin-1/2 particle that interacts with the  $SU(3)$  color force. Just as with leptons, there are six flavors of quarks, arranged in three families or generations. Following very closely what we did with the leptons, we work with only one generation. Extending to the other generators is then trivial. To begin, define three fields:  $q$ ,  $\bar{u}$ , and  $\bar{d}$ , in the representations  $(3, 2, 1/6)$ ,  $(\bar{3}, 1, -2/3)$ , and  $(\bar{3}, 1, 1/2)$  of  $SU(3) \times SU(2) \times U(1)$ . The field  $q$  will be the

$SU(2)$  doublet

$$q = \begin{pmatrix} u \\ d \end{pmatrix}. \quad (25.43)$$

This is exactly analogous to equation (25.28).

Again, following what we did with the leptons, we can write out the covariant derivative for all three fields:

$$\begin{aligned} (D_\mu q)_{\alpha i} &= \partial_\mu q_{\alpha i} - ig_3 W_\mu^a (T_2^a)^\beta_\alpha q_{\beta i} - ig_2 W_\mu^a (T_2^a)^j_i q_{\beta j} - ig_1 \left(\frac{1}{6}\right) B_\mu q_{\alpha i}, \\ (D_\mu \bar{u})^\alpha &= \partial_\mu \bar{u}^\alpha - ig_3 W_\mu^a (T_3^a)^\alpha_\beta \bar{u}^\beta - ig_1 \left(-\frac{2}{3}\right) B_\mu \bar{u}^\alpha, \\ (D_\mu \bar{d})^\alpha &= \partial_\mu \bar{d}^\alpha - ig_3 W_\mu^a (T_3^a)^\alpha_\beta \bar{d}^\beta - ig_1 \left(\frac{1}{3}\right) B_\mu \bar{d}^\alpha, \end{aligned} \quad (25.44)$$

where  $i$  is an  $SU(2)$  index and  $\alpha$  is an  $SU(3)$  index. The  $SU(3)$  index is lowered for the  $\mathbf{3}$  representation and raised for the  $\bar{\mathbf{3}}$  representation.

Using the non-Cartan and Cartan parts of the covariant derivatives in terms of the lower energy  $SU(2) \times U(1)$  gauge fields,

$$\begin{aligned} g_2 W_\mu^1 T^1 + g_2 W_\mu^2 T^2 &= \frac{g_2}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}, \\ g_2 W_\mu^3 T^3 + g_1 B_\mu Y &= eQ A_\mu + \frac{e}{s_w c_w} (T^3 - s_w^2 Q) Z_\mu, \end{aligned} \quad (25.45)$$

it is again straightforward to find the electric charge eigenvalue for each quark field:

$$Qu = +\frac{2}{3}u, \quad Qd = -\frac{1}{3}, \quad Q\bar{u} = -\frac{2}{3}\bar{u}, \quad Q\bar{d} = +\frac{1}{3}\bar{d}. \quad (25.46)$$

The primary idea to take away is that the  $SU(2)$  doublet (25.43) behaves exactly as the lepton doublet in (25.28) when interacting with the 'raising' and 'lowering' gauge particles  $W_\mu^\pm$ . This is why the  $u$  and  $d$  are arranged in the  $SU(2)$  doublet  $q$  in (25.43), and why  $q$  carries the  $SU(2)$  index  $i$  in the covariant derivative (25.44), whereas  $\bar{u}$  and  $\bar{d}$  carry only the  $SU(3)$  index.

The  $SU(3)$  index runs from 1 to 3, and the 3 values are conventionally denoted red, green, and blue ( $r, g, b$ ). These obviously are merely labels and have nothing to do with the colors in the visible spectrum.

The eight gauge fields associated with the eight  $SU(3)$  generators are called *Gluons*, and they are represented by the matrices. We label each gluon as

follows:

$$g_{\alpha}^{\beta} = \begin{pmatrix} r\bar{r} & r\bar{g} & r\bar{b} \\ g\bar{r} & g\bar{g} & g\bar{b} \\ b\bar{r} & b\bar{g} & b\bar{b} \end{pmatrix}, \quad (25.47)$$

so that the upper index is the anti-color index, and denotes the column of the matrix, and the lower index is the color index denoting the row of the matrix. Then consider the gluon

$$g_r^{\bar{g}} \propto \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (25.48)$$

and the quarks

$$q_r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad q_g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad q_b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (25.49)$$

It is easy to see that this gluon will interact as

$$g_r^{\bar{g}} q_r = 0, \quad g_r^{\bar{g}} q_g = q_r, \quad g_r^{\bar{g}} q_b = 0. \quad (25.50)$$

Or in other words, the gluon with the anti-green index will only interact with a green quark. There will be no interaction with the other quarks. Multiplying this out, and looking more closely at the behavior of the  $SU(3)$  generators and eigentstates, you can work out all of the interaction rules between quarks and gluons. You will see that they behave exactly according to the root space of  $SU(3)$ .

## 25.4 The Yukawa Couplings

We already has discussed the idea of renormalization. We said that certain theories can be renormalized and others cannot. It turns out that while the theory we have outlined so far is renormalizable, if we try to add mass terms for  $l$  (left lepton doublets) and  $\bar{e}$  (right anti-singlets) fields, the theory breaks down. Therefore we cannot add a mass term. But, we know experimentally that electrons and neutrinos have mass. To incorporate mass into the theory we must use the Higgs mechanism by adding of a Yukawa term,

$$\mathcal{L}_{\text{Yukawa}}^{\text{Leptons}} = -y\epsilon^{ij}\phi_i l_j \bar{e} + \text{h.c.}, \quad (25.51)$$

where  $y$  is another coupling constant,  $\epsilon^{ij}$  is the totally antisymmetric tensor, and h.c. is the Hermitian conjugate of the first term.

Now that we have added  $\mathcal{L}_{\text{Yukawa}}$  to the Lagrangian, we want to break the symmetry exactly as we did in the previous lecture. First, we replace  $\phi_1$  with  $(v+h)/\sqrt{2}$  and  $\phi_2$  with 0. So,

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}}^{\text{Leptons}} &= -y\epsilon^{ij}\phi_i l_j \bar{e} + \text{h.c.} = -y(\phi_1 l_2 - \phi_2 l_1)\bar{e} + \text{h.c.} = \\ &= -\frac{1}{\sqrt{2}}y(v+h)l_2\bar{e} + \text{h.c.} = \\ &= -\frac{1}{\sqrt{2}}y(v+h)(e\bar{e} + \bar{e}e) = -\frac{1}{\sqrt{2}}y(v+h)\bar{\mathcal{E}}\mathcal{E} ,\end{aligned}\tag{25.52}$$

where

$$\mathcal{E} = \begin{pmatrix} e \\ \bar{e}^\dagger \end{pmatrix}\tag{25.53}$$

is the Dirac field for the electron ( $e$  is the electron and  $\bar{e}^\dagger$  is the antielectron, or positron). We see that in (25.52) it is a mass term for the electron and positron.

Just as with leptons, we cannot write down a mass term for quarks, but we can include a Yukawa term coupling these fields to the Higgs:

$$\mathcal{L}_{\text{Yukawa}}^{\text{Quarks}} = -y'\epsilon^{ij}\phi_i q_{\alpha j} \bar{d}^\alpha - y''\phi^\dagger{}^i q_{\alpha i} \bar{u}^\alpha + \text{h.c.} .\tag{25.54}$$

As with the leptons, we can break the symmetry and writing out this Yukawa term, we get

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}}^{\text{Quarks}} &= -\frac{1}{\sqrt{2}}(v+h) \left[ y'(d_\alpha \bar{d}^\alpha + \bar{d}_\alpha^\dagger d^{\dagger\alpha}) + y''(u_\alpha \bar{u}^\alpha + \bar{u}_\alpha^\dagger u^{\dagger\alpha}) \right] = \\ &= -\frac{1}{\sqrt{2}}(v+h) (y'\bar{\mathcal{D}}^\alpha \mathcal{D}^\alpha + y''\bar{\mathcal{U}}^\alpha \mathcal{U}_\alpha) ,\end{aligned}\tag{25.55}$$

where we have defined the Dirac fields for the up and down quarks:

$$\mathcal{D}_\alpha \equiv \begin{pmatrix} d_\alpha \\ \bar{d}_\alpha^\dagger \end{pmatrix} , \quad \mathcal{U}_\alpha \equiv \begin{pmatrix} u_\alpha \\ \bar{u}_\alpha^\dagger \end{pmatrix} .\tag{25.56}$$

Notice that, whereas both the up and down quarks were massless before breaking, they have now acquired masses

$$m_d = \frac{y'v}{\sqrt{2}} , \quad m_u = \frac{y''v}{\sqrt{2}} .\tag{25.57}$$

To resume, the last term in the EW Lagrangian (25.1) can be written in the form:

$$\begin{aligned}
 \mathcal{L}_{\text{Yukawa}} &= \mathcal{L}_{\text{Yukawa}}^{\text{Leptons}} + \mathcal{L}_{\text{Yukawa}}^{\text{Quarks}} = \\
 &= - \sum_{m,n=1}^F \left[ \Gamma_{mn}^u \bar{q}_L^m \bar{\phi} u_R^n + \Gamma_{mn}^d \bar{q}_L^m \phi d_R^n + \right. \\
 &\quad \left. + \Gamma_{mn}^e \bar{l}_L^m \phi e_R^n + \Gamma_{mn}^\nu \bar{l}_L^m \bar{\phi} \nu_R^n \right] + h.c. , \tag{25.58}
 \end{aligned}$$

where the matrices  $\Gamma_{mn}$  describe the Yukawa couplings between the single Higgs doublet,  $\phi$ , and the various flavors  $m$  and  $n$  of quarks and leptons. One needs representations of Higgs fields with the hyper-charge  $+1/2$  and  $-1/2$  to give masses to the down quarks and electrons ( $+1/2$ ), and to the up quarks and neutrinos ( $-1/2$ ). The representation  $\phi^\dagger$  has the hyper-charge  $-1/2$ , but transforms as the  $\bar{2}$  rather than the  $2$ . However, in  $SU(2)$  the  $\bar{2}$  representation is related to the  $2$  by a similarity transformation, and

$$\bar{\phi} \equiv i\tau^2 \phi^\dagger = (\phi^{0\dagger}, -\phi^-) \tag{25.59}$$

transforms as a  $2$ . All of the masses can therefore be generated with a single Higgs doublet if one makes use of both  $\phi$  and  $\bar{\phi}$ . The fact that the fundamental and its conjugate are equivalent does not generalize to higher unitary groups.

**Exercise 25.1:** Find the gauge boson propagator in momentum space in axial and Coulomb gauges.

**Exercise 25.2:** Find the transformation which brings the equation for a complex scalar field  $\phi$  in the electromagnetic field of the form  $A^\nu = (A^0, 0)$  to the equation of  $\phi^*$ .

**Exercise 25.3:** Draw all possible lowest-order Feynman diagrams for the processes:  $e^+e^- \rightarrow \mu^+\mu^-$ ;  $e^+e^- \rightarrow \nu_\mu \bar{\nu}_\mu$ ;  $\nu_\mu e^- \rightarrow \nu_\mu e^-$  and  $\bar{\nu}_e e^- \rightarrow \bar{\nu}_e e^-$ .

**Exercise 25.4:** Explain why the decay  $D^0 \rightarrow K^-\pi^+$  is possible, while  $D^0 \rightarrow K^+\pi^-$  is not allowed.



# Chapter 26

## Interactions

Now let us write out in details interacting terms of gauge bosons and fermions within the SM.

### 26.1 Charged Currents

In the EW model the interaction of the  $W_\mu^\pm$  bosons to fermions is given by

$$\mathcal{L} = -\frac{g_2}{2\sqrt{2}} \left( J_W^\mu W_\mu^- + J_W^{\mu\dagger} W_\mu^+ \right), \quad (26.1)$$

where the weak charge-raising current is

$$\begin{aligned} J_W^{\mu\dagger} &= \sum_{m=1}^N [\bar{\nu}_m \gamma^\mu (1 - \gamma^5) e_m + \bar{u}_m \gamma^\mu (1 - \gamma^5) d_m] = \\ &= (\bar{\nu}_e \bar{\nu}_\mu \bar{\nu}_\tau) \gamma^\mu (1 - \gamma^5) V_\ell \begin{pmatrix} e^- \\ \mu^- \\ \tau^- \end{pmatrix} + (\bar{u} \bar{c} \bar{t}) \gamma^\mu (1 - \gamma^5) V_q \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \end{aligned} \quad (26.2)$$

$J_W^{\mu\dagger}$  has a  $V - A$  form, i.e. it violates parity and charge conjugation maximally. The vector and axial fermion gauge interaction vertices for  $\bar{d}_j u_i W^-$  vertex is the same as for  $\bar{u}_i d_j W^+$  except

$$V_{qij} \rightarrow \left( V_q^\dagger \right)_{ji} = V_{qij}^*. \quad (26.3)$$

The lepton- $W^\pm$  vertices are obtained from the quark ones by  $u_i \rightarrow \nu_i$ ,  $d_j \rightarrow e_j^-$  and  $V_q \rightarrow V_\ell$ .

The mismatch between the unitary transformations relating the weak and mass eigenstates for the up and down-type quarks leads to the presence of the  $N \times N$  unitary matrix  $V_q \equiv A_L^{u\dagger} A_L^d$  in the CC. This is the Cabibbo-Kobayashi-Maskawa matrix, which is ultimately due to the mismatch between the weak and Yukawa interactions. For  $N = 2$  families  $V_q$  takes the familiar form

$$V_{\text{Cabibbo}} = \begin{pmatrix} \cos \theta_c & \sin \theta_c \\ -\sin \theta_c & \cos \theta_c \end{pmatrix}, \quad (26.4)$$

where  $\sin \theta_c \simeq 0.22$  is the Cabibbo angle. This form gives a good zero<sup>th</sup>-order approximation to the weak interactions of the  $u$ ,  $d$ ,  $s$  and  $c$  quarks; their coupling to the third family, though non-zero, is very small. Including these couplings, the 3-family Kobayashi-Maskawa matrix is

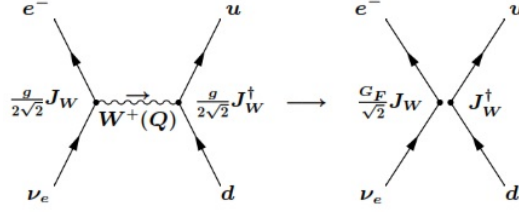
$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \sim \begin{pmatrix} 1 & \lambda & \lambda^3 \\ \lambda & 1 & \lambda^2 \\ \lambda^3 & \lambda^2 & 1 \end{pmatrix}, \quad (26.5)$$

where the  $V_{ij}$  may involve a  $CP$ -violating phase. The second form, with  $\lambda = \sin \theta_c$  is an easy to remember approximation to the observed magnitude of each element, which displays a suggestive but not well understood hierarchical structure. These are order of magnitude only; each element may be multiplied by a phase and a coefficient of  $\mathcal{O}(1)$ .

As we know, an arbitrary  $N \times N$  unitary matrix involves  $N^2$  real parameters. In this case  $2N - 1$  of them are unobservable relative phases in the fermion mass eigenstate fields, leaving  $N(N - 1)/2$  rotation angles and  $(N - 1)(N - 2)/2$  observable  $CP$ -violating phases. There are an additional  $N - 1$  Majorana phases in  $V_\ell$  for Majorana neutrinos.

In (26.2)  $V_\ell \equiv A_L^{\nu\dagger} A_L^e$  is the analogous leptonic mixing matrix. It is critical for describing neutrino oscillations and other processes sensitive to neutrino masses. However, for processes for which the neutrino masses are negligible we can effectively set  $V_\ell = I$ . More precisely,  $V_\ell$  will only enter such processes in the combination  $V_\ell^\dagger V_\ell = I$ , so it can be ignored.

In interactions mediated by the exchange of a  $W_\mu^\pm$ :



in the limit  $|Q^2| \ll M_W^2$  the momentum term in the  $W$ -propagator can be neglected. This leads to an effective zero-range (four-fermi) interaction

$$-\mathcal{L}_{eff}^{cc} = \frac{G_F}{\sqrt{2}} J_W^\mu J_{W\mu}^\dagger, \quad (26.6)$$

where the Fermi constant is identified as

$$\frac{G_F}{\sqrt{2}} \simeq \frac{g_2^2}{8M_W^2} = \frac{1}{2v^2}. \quad (26.7)$$

Thus, the Fermi theory is an approximation to the SM valid in the limit of small momentum transfer.

The weak CC interaction as described by (26.6) has been successfully tested in a large variety of weak decays, including  $\beta$ ,  $K$ , hyperon, heavy quark,  $\mu$ , and  $\tau$  decays. Weak CC effects have also been observed in higher orders, such as in  $K^0 - \bar{K}^0$ ,  $D^0 - \bar{D}^0$ , and  $B^0 - \bar{B}^0$  mixing, and in  $CP$  violation in  $K$  and  $B$  decays. For these higher order processes the full theory must be used because large momenta occur within the loop integrals.

### 26.1.1 QED Part

The SM incorporates all of the successes of QED, which is based on the  $U(1)_{em}$  subgroup that remains unbroken after spontaneous symmetry breaking. The relevant part of the Lagrangian density is

$$\mathcal{L} = -\frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} J_Q^\mu (\cos \theta_W B_\mu + \sin \theta_W W_\mu^3), \quad (26.8)$$

where the linear combination of neutral gauge fields is just the photon field  $A_\mu$ . This reproduces the QED interaction provided one identifies the combination of couplings  $e = g_2 \sin \theta_W$  as the electric charge of the positron,

where  $\tan \theta_W \equiv g_1/g_2$ . The electromagnetic current is given by

$$J_Q^\mu = \sum_{m=1}^N \left[ \frac{2}{3} \bar{u}_m \gamma^\mu u_m - \frac{1}{3} \bar{d}_m \gamma^\mu d_m - \bar{e}_m \gamma^\mu e_m \right]. \quad (26.9)$$

It takes the same form when written in terms of either weak or mass eigenstates because all fermions which mix with each other have the same electric charge. Thus, the electromagnetic current is automatically flavor-diagonal.

QED is the most successful theory in physics when judged in terms of the theoretical and experimental precision of its tests. The approximate agreement of these determinations, which involves the calculation of the electron anomalous magnetic moment  $a_e = (g_e - 2)/2$  to high order, validates not only QED but the entire formalism of gauge invariance and renormalization theory. Other basic predictions of gauge invariance (assuming it is not spontaneously broken, which would lead to electric charge non-conservation), are that the photon mass  $m_\gamma$  and its charge  $q_\gamma$  (in units of  $e$ ) should vanish. The current upper bounds are extremely impressive  $m_\gamma < 1 \times 10^{-18}$  eV,  $q_\gamma < 5 \times 10^{-30}$ , based on astrophysical effects (the survival of the Solar magnetic field and limits on the dispersion of light from pulsars).

## 26.2 Neutral Currents

The third class of gauge interactions is the weak NC, which was predicted by the  $SU(2) \times U(1)$  model. The relevant interaction is

$$\mathcal{L} = -\frac{\sqrt{g_1^2 + g_2^2}}{2} J_Z^\mu (-\sin \theta_W B_\mu + \cos \theta_W W_\mu^3) = -\frac{g_2}{2 \cos \theta_W} J_Z^\mu Z_\mu, \quad (26.10)$$

where the combination of neutral fields is the massive  $Z$ -boson field, the value of which strength in is conveniently rewritten using

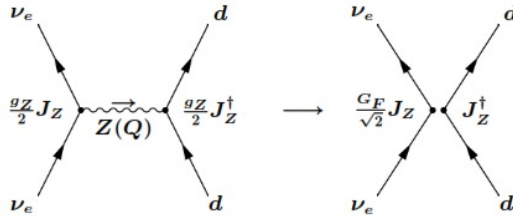
$$\cos \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (26.11)$$

The weak NC is given by

$$J_Z^\mu = \sum_m \left[ \bar{u}_{mL} \gamma^\mu u_L^m - \bar{d}_{mL} \gamma^\mu d_L^m + \bar{\nu}_{mL} \gamma^\mu \nu_L^m - \bar{e}_{mL} \gamma^\mu e_L^m \right] - 2 \sin^2 \theta_W J_Q^\mu. \quad (26.12)$$

Like the electromagnetic current  $J_Z^\mu$  is flavor-diagonal in the SM; all fermions which have the same electric charge and chirality and therefore can mix with each other have the same  $SU(2) \times U(1)$  assignments, so the form is not affected by the unitary transformations that relate the mass and weak bases. It was for this reason that the GIM mechanism was introduced into the model, along with its prediction of the charm quark. Without it the  $d$  and  $s$  quarks would not have had the same  $SU(2) \times U(1)$  assignments, and flavor-changing neutral currents would have resulted. The absence of such effects is a major restriction on many extensions of the SM involving exotic fermions.

The NC has two contributions. The first only involves the left-chiral fields and is purely  $V - A$ . The second is proportional to the electromagnetic current with coefficient  $\sin^2 \theta_W$  and is purely vector. Parity is therefore violated in the neutral current interaction, though not maximally.



On the Figure above you can see the typical neutral current interaction mediated by the exchange of the  $Z$ , which reduces to an effective four-fermi interaction in the limit that the momentum transfer  $Q$  can be neglected.  $g_Z$  is defined as  $\sqrt{g_1^2 + g_2^2}$ .

In an interaction between fermions in the limit that the momentum transfer is small compared to  $M_Z$  one can neglect the  $Q^2$  term in the propagator, and the interaction reduces to an effective four-fermi interaction

$$-\mathcal{L}_{eff}^{NC} = \frac{G_F}{\sqrt{2}} J_Z^\mu J_{Z\mu} . \quad (26.13)$$

The coefficient is the same as in the charged case because

$$\frac{G_F}{\sqrt{2}} = \frac{g_2^2}{8M_W^2} = \frac{g_1^2 + g_2^2}{8M_Z^2} . \quad (26.14)$$

That is, the difference in  $Z$ -couplings compensates the difference in masses in the propagator.

The weak NC was discovered at CERN in 1973 and at Fermilab shortly thereafter, and since that time  $Z$  exchange and  $\gamma - Z$  interference processes have been extensively studied in many interactions, including  $\nu e \rightarrow \nu e$ ,  $\nu N \rightarrow \nu N$  and  $\nu N \rightarrow \nu X$  polarized  $e^-$ -hadron and  $\mu$ -hadron scattering; atomic parity violation; and in  $e^+e^-$  and  $Z$ -pole reactions. Along with the properties of the  $W$  and  $Z$  they have been the primary quantitative test of the unification part of the EW model. The results of these experiments have been in excellent agreement with the predictions of the SM, indicating that the basic structure is correct to first approximation and constraining the effects of possible new physics.

### 26.3 Anomaly Cancellation

As we have seen above, the EW model contains both the vector and the axial vector currents. This causes a problem when we try to renormalize the theory and do loop computations. The gauge invariance of axial vector currents of the type

$$J_\mu^5 = \bar{\psi} \gamma_\mu \gamma_5 \psi' \quad (26.15)$$

( $\psi' = \psi$  for neutral currents), is not preserved by dimensional regularization due to the presence of  $\gamma_5$  in the current. This means that even though,  $\partial_\mu J_5^\mu = 0$  classically, at loop level due to the non-invariance of the regulator,  $\partial_\mu J_5^\mu \neq 0$  and develops a nonzero term on the right hand side. Hence, this axial gauge current is no longer conserved. The current is said to be ‘anomalous’.

As we know from Noether’s theorem if the current is not conserved, it means gauge invariance is broken. Gauge symmetry along with Higgs mechanism is needed to have a consistent quantum theory with massive gauge bosons. Thus if the theory has an anomalous current (or has anomaly) the theory may not make sense at quantum level. It was shown by Adler and Bell-Jakciw, that there is only one type of loop diagram with a logarithmic divergence which can make  $\partial_\mu J_5^\mu$  non-vanishing and poses a danger to the conservation of the axial gauge current. This is a triangle diagram with a fermion loop and two gauge boson legs and one current insertion; equivalently one can also consider a fermion loop with three gauge boson legs. In the EW model with its  $SU(2)_L$  gauge bosons which have couplings only to left chiral fermions and the  $U(1)_Y$  gauge bosons which have unequal couplings to the left and right chiral fermions, these triangle diagrams are in general not zero. Further, one

can show that the anomalous contribution is independent of the mass of the fermions in the internal loop.

There are in fact four types of triangle diagrams we need to consider out of which three are shown in the Figure below.

Consider the diagram in the left most panel which contains matrix element of a pure  $V - A$  current insertion along with two  $SU(2)_L$  gauge boson legs. Only left handed fermions contribute to this anomaly and it can be shown that

$$\partial_\mu J_5^a \sim \text{tr} \tau^a \{ \tau^b, \tau^c \} \epsilon^{\alpha\rho\beta\nu} F_{\alpha\rho}^b F_{\beta\nu}^c . \tag{26.16}$$

Here the  $\text{tr}$  refers to the trace over representation matrices and indicates the sum over all the fermions in the representation. Since  $\{ \tau^b, \tau^c \} = 2\delta^{bc}$  and  $\tau^a$  are traceless matrices this anomaly is zero identically. In fact, the diagram with just one  $SU(2)_L$   $V - A$  current insertion not shown here will also give zero contribution to the anomaly due to the traceless property of  $\tau^a$  ( $a = 1, 2, 3$ ) matrices.

The central diagram also gets contribution only from the left chiral fermions and is given by

$$\partial_\mu J_5^\mu \sim \text{tr}(Y_L) \epsilon^{\alpha\rho\beta\nu} F_{\alpha\rho}^a F_{\beta\nu}^a . \tag{26.17}$$

The notation  $\text{tr}(Y_L)$  indicates that only the left chiral fermions contribute to this quantity and sum is to be taken over one  $SU(2)_L$  representation.

The contribution of the rightmost diagram is given by:

$$\partial_\mu J_5^\mu \sim \text{tr} (Y_L^3 - Y_R^3) \epsilon^{\alpha\rho\beta\nu} B_{\alpha\rho} B_{\beta\nu} . \tag{26.18}$$

We see that for a single lepton generation the anomaly of (26.17) is proportional to  $2 \times Y_L = -2$ . Summing over all the lepton doublets it will have a value  $-6$ . However, one notices that, for a single quark generation it is  $2 \times 1/3$ . The three colors add another factor of 3. Thus we find

$\text{tr}(Y_L)|_l + \text{tr}(Y_L)|_q = -2 + 3 \times 2 \times 1/3 = 0$ . Thus this anomaly vanishes identically for the particle content of the left chiral fermions in the EW model. Further, we also notice that while  $(2Y_L^l)^3 - (Y_R^e)^3 = -2 + 8 = 6$  is not zero, it is again compensated by the value for the quark doublets which is

$$3 \times \left[ -\frac{2}{27} - \left(\frac{4}{3}\right)^3 + \left(\frac{2}{3}\right)^3 \right] = -6. \quad (26.19)$$

Thus again  $\text{tr}(Y_L^3 - Y_R^3)|_l + \text{tr}(Y_L^3 - Y_R^3)|_q = 6 - 6 = 0$ . Hence contributions to both the anomalies, from loops of fermions of one quark and one lepton doublet of the EW model, are equal and opposite in sign. This means that the numbers of the lepton and quark doublets have to be exactly equal so that the anomalies do not spoil the gauge invariance of the EW model and hence the renormalizability.

To conclude, the SM is anomaly free for the assumed fermion content. There are no  $SU(3)^3$  anomalies because the quark assignment is non-chiral, and no  $SU(2)^3$  anomalies because the representations are real. The  $SU(2)^2 Y$  and  $Y^3$  anomalies cancel between the quarks and leptons in each family, by what appears to be an accident. The  $SU(3)^2 Y$  and  $Y$  anomalies cancel between the  $L$  and  $R$  fields, ultimately because the hypercharge assignments are made in such a way that  $U(1)_Q$  will be non-chiral.

**Exercise 26.1:** Find an expression for the  $\nu_\mu e^-$  NC cross section in terms of the laboratory frame neutrino energy.

**Exercise 26.2:** Write out vertex factors for the five following interactions:  $d_L + W^+ \rightarrow u_L$ ;  $d_L + Z \rightarrow d_L$ ;  $u_L + Z \rightarrow u_L$ ;  $u_L + \gamma \rightarrow u_L$  and  $u_{Lr} + g_{rb} \rightarrow u_{Lb}$ .

**Exercise 26.3:** A beam of unpolarized electrons:

- Can be described by a wave function that is an equal superposition of spin-up and spin-down wave functions;
- Cannot be described by a wave function;
- Neither of the above.

**Exercise 26.4:** Assume that the same basic weak interaction is responsible for the beta decay processes  $n \rightarrow pe^- \bar{\nu}$  and  $\Sigma \rightarrow \Lambda e^- \bar{\nu}$ , and that the matrix elements describing these decays are the same. Estimate the decay rate of the  $\Sigma$ -decay, given the lifetime of a free neutron is about  $10^3$  seconds.



## Chapter 27

# Predictions of the SM

Let us summarize some of the qualitative and quantitative implications of the  $SU(2)_L \times U(1)_Y$  invariance. Note that almost all of them are result of the invariance and hence not specific to the actual mechanism of symmetry breaking as long as it preserves the symmetry.

1. First and foremost, this is a unification of weak and electromagnetic interaction: i.e.  $e$ ,  $g_1$ ,  $g_2$  all are of similar order and the apparent difference in strengths of electromagnetic interactions ( $\alpha_{\text{em}}$  and  $G_\mu$ ), is only caused by the large value of the masses of the weak gauge bosons compared to that of the massless photon. The model predicts existence of a new weak gauge boson  $Z_\mu$  and that of the weak NC mediated by it, analogous to the weak CC mediated by the  $W_\mu^\pm$ . Further, the strength of this new weak interaction is similar to that of the CC weak interaction. This is particularly transparent once we use the  $\rho = 1$  prediction (25.24) of the EW model wherein  $W/Z$  masses are generated by SSB using a Higgs doublet.
2. Further, flavor changing NCs are absent at tree level if and only if all the quarks of a given electrical charge belong to the same representation of  $SU(2)_L$ . Thus the experimentally observed absence of flavor changing NC requires existence of the charm quark  $c$ , in addition to the already known  $u, d$  and  $s$  quarks and one could also ‘predict’ the mass of the  $c$  quark from the measured  $K_0-\bar{K}_0$  mass difference.
3. Since  $G_\mu$  and the electron charge  $e$  are measured experimentally, the

model has two free parameters,  $\sin\theta_W$  and  $M_W$ . If one assumes  $g_2 = e$ , i.e.  $\sin\theta_W = 1$ , then we get  $M_W \sim \mathcal{O}(100)GeV$ . However, when the gauge boson masses are generated through the SSB,  $M_W$  can be expressed in terms of  $G_\mu, \alpha_{em}$  and  $\sin^2\theta_W$ .

4. The model predicts precise nature of the  $WWZ$  coupling, the strength being given by  $g_2$ .
5. Couplings of all the fermions with the gauge boson  $Z_\nu$  are determined in terms of  $\sin\theta_W$ , once the representations of the two gauge groups to which the fermions belong are specified.
6. Requirement of anomaly cancellation, necessary for the renormalizability, predicts that the number of lepton and quark generation seen in nature should be equal. So while the model cannot predict how many families of quarks and leptons there should be, it predicts their equality.
7. The conditions of anomaly cancellation and observed closeness of  $\rho$  to unity, then gives strong constraints on new particles that one can be added to the spectrum of the EW model.
8. As already stated in the previous lecture, generation of gauge bosons masses via SSB provides some more relations among physical quantities and hence reduces the number of free parameters of the model to one, that parameter being  $\sin\theta_W$ .

Thus this model could be easily subjected to experimental tests. This is what we will discuss in this lecture.

## 27.1 High Energy Scattering

We saw how the postulate of massive vector boson was inspired by the demand to restore unitarity to the neutrino induced processes. For example, the amplitude (say) for  $\nu e \rightarrow \nu e$  scattering calculated in Fermi theory (current-current interactions) violates tree level unitarity for  $\sqrt{s} \sim 300 \sim G_\mu^{-1/2}$  GeV. Hence, one could also take this value as an upper bound on the mass of the 'massive'  $W_\mu^\pm$  boson.

However, theories with massive vector bosons have problems with gauge invariance and hence renormalizability. The SSB via Higgs mechanism solved

the problem by generating these masses in a gauge invariant manner. This then meant that the theory has renormalizability even with massive gauge bosons. In fact, as we will discuss below, we can see explicitly that gauge invariance also renders nice high energy behavior to all the scattering amplitudes of the EW theory.

The existence of massive vector gauge bosons restore unitary behavior to processes like  $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$ . But now due to the same non zero mass of the  $W_\mu^\pm$  bosons, amplitudes for processes involving longitudinal  $W$ 's have a bad high energy behaviour. For example, the matrix element for the process  $\nu_e \bar{\nu}_e \rightarrow W^+ W^-$  through a  $t$ -channel exchange of an  $e$ , shown in the left panel of the Figure below,

grows too fast with energy and violates unitarity. One can show that

$$\mathcal{M}(\nu_e \bar{\nu}_e \rightarrow W^+ W^-) \sim 8 \frac{g_2^2}{M_W^2} E p' \sin \theta, \quad (27.1)$$

where  $E$  is the energy of the incoming  $\nu_e$  and  $p', \theta$  are the momentum and the angle of scattering of the  $W$  boson in the final state. Here we write only the dominant term of the amplitude involving the longitudinal gauge bosons, which is the one with bad high energy behavior. If one does a partial wave analysis of this amplitude, one finds that this amplitude will violate partial wave unitarity, for  $s \sim M_W^2 / 2g_2^2$ . However, what is interesting is that the contribution to the matrix element of the process  $\nu_e \bar{\nu}_e \rightarrow W^+ W^-$ , from the  $s$  channel exchange of a  $Z$  boson, shown in the right panel of the Figure above has exactly the same magnitude as the  $t$  channel contribution written  $\bar{\nu}_e$  above but opposite in sign. This happens only if the strength and structure of the couplings of the  $Z$  with a  $\nu$  and  $W$  pair is exactly the same as given by

the  $SU(2)_L \times U(1)$  theory. Thus the violation of unitarity in the amplitude  $\nu_e \bar{\nu}_e \rightarrow W^+ W^-$  due to the longitudinal gauge boson scattering is cured in a gauge theory.

In fact, the EW model contains more such amplitudes which, in principle, could have had bad high energy behavior but which are rendered safe by the particle content and the coupling structure of the SM. It was demonstrated that in the EW model where the masses are generated through SSB by a Higgs doublet, all such amplitudes satisfy tree level unitarity. In fact the leading divergence of the  $\mathcal{M}(WW \rightarrow WW)$  which goes like  $s^2$  and hence is much worse, is also cured by the  $Z$  exchange contribution and the contribution of the quartic coupling among the  $W$  bosons which arise from the non-Abelian gauge invariance of the theory. Further, the divergent term proportional to  $s$  is cancelled by the contribution of the process  $W^+ W^- \rightarrow h \rightarrow W^+ W^-$ , where the Higgs boson is exchanged in the  $s$ -channel. Also if one were to calculate high energy behavior of the amplitude of the process  $e^+ e^- \rightarrow W^+ W^-$ :

then the same cancellation between the divergent parts of the  $t$ -channel and  $s$ -channel amplitudes is seen to take place.

After this observation, a variety of authors investigated the conditions necessary for cancellation of these divergences so that the amplitudes will satisfy tree level unitarity. In fact their analysis indicated that this requires existence of partial wave contributions in the spin 1 and spin 0 channel, with the couplings of these particles exchanged in the  $s$ -channel to be precisely those that are given the SM. Recall here that this proportionality of the coupling of the Higgs to the masses of the particles to which it couples is the key prediction of the SSB by Higgs mechanism. The other couplings are of course given by the gauge invariance itself. Thus one could have derived the existence of the Higgs boson as well as the structure of the couplings of the fermions and the gauge bosons to it, without making any reference to the Higgs mechanism and hence the renormalizability.

The fact that the two different requirements, unitarity and renormalizability, lead us to the same result, indicates that there must be a deep connection between the two. In fact, for the  $\nu_e\bar{\nu}_e \rightarrow W^+W^-$  scattering, there is a residual logarithmic violation of unitarity that is left after all the cancellations, which gets cancelled by the scale dependence of  $g_2$  which is a loop effect which can be computed reliably only in a renormalizable theory.

## 27.2 Observations Meet Predictions

To start, let us just briefly take a look how the establishment of the SM has been a synergistic activity between theoretical and experimental developments. We saw already how the form of the pre-gauge theory, effective Hamiltonian description of weak interactions, obtained phenomenologically from the data hinted at a possible gauge theoretic description of the same. Equally interesting are the hints at existence of new particles given by the theory. While some of the members of this periodic table, like the  $\mu$ , were unlooked for and some like the  $\nu$  were met with quite a bit of disbelief when postulated theoretically, for most of the recent additions their existence and in some cases even their masses were predicted if the EW interactions were to be described by a renormalizable gauge theory.

In fact, the existence of strange particles which contain the strange quarks, coupled with experimental features such as the suppression of the flavor changing NC in EW processes alluded to before, indicated the existence of the charm quark, as already indicated above. Further, the small mass difference between  $K_L$  and  $K_S$  (or alternatively the  $K_0-\bar{K}_0$  mixing) could be used to obtain an estimate of its mass. Accidental discovery of some members of the third lepton and quark family, combined with the requirement of anomaly cancellation, an essential feature for a renormalizable theory, meant that the remaining members of the same family had to exist. Hence  $t$  and the  $\nu_\tau$  were hunted for very actively once the  $b$  and the  $\tau$  made their appearance! The properties of a renormalizable quantum field theory were the essential reasons behind the belief in these predictions. The mass of the  $t$  quark could also be predicted in the SM, using experimental information on neutral B meson mixing and properties of the  $Z$  boson, as we will see below.

The story is not very different for the EW gauge bosons. As was already mentioned, requiring consistency of the pre gauge theory description of the weak

interactions with unitarity, had indicated a nonzero mass for the charged  $W^\pm$  but had not indicated what the mass would be, except that it should be much larger than the typical energy scales involved in the weak decays. It is the unified description of the EW interactions that actually gave a lower limit on its mass. Note that the correctness of the  $V - A$  nature of weak interactions and pure vector nature of the electromagnetic interactions predicted existence of a neutral boson other than the photon. The masses of the  $W$  and the new  $Z$  boson required in the unified EW theory, were all predicted in terms of the life time of the  $\mu$  and the weak mixing angle  $\theta_W$  which was a free parameter in the model. This could be determined from measurements of rates of various weak processes.

Not just this, the SM also had predicted existence of yet another boson, this time with spin 0 – the Higgs boson. The mass of the Higgs boson, however, is a free parameter in the framework of the SM. Comparisons of the EW observables with precision measurements can constrain the Higgs mass through the corrections caused by the loop effects which can be computed in a renormalizable quantum field theory. One can also put limits on this parameter from theoretical considerations of consistency of the SM as a field theory at high scales: the triviality and vacuum stability, discussed in the previous lecture.

Let us discuss in detail the case of the  $t$  quark which is quite interesting. The existence of the  $t$  quark and the information on its mass came from a variety of theoretical and phenomenological observations in flavor physics and physics of the  $W/Z$  bosons. As already mentioned the explanation of the experimentally observed CP violation in terms of the quark mixing matrix requires at least three generations of quarks. This mixing is described by the Kobayashi-Maskawa mixing matrix. So in that sense existence of the  $t$  and  $b$  was indicated by this observation. The requirement of anomaly cancellation for the gauge theory of EW interactions to be renormalizable, further indicated existence of an additional generation of leptons,  $\tau$  and  $\nu_\tau$  as well. Experimental manifestation of  $B_0-\bar{B}_0$  oscillations was a harbinger of the presence of the  $t$  quark. Further indications for the expected mass actually came from precision measurements of many EW observables, i.e. properties of the  $Z$  and the  $W$  boson and the quantum corrections caused to them by loops containing top quarks.

Experimental observation of the  $t$  quark, with a mass value consistent with the implications of the EW precision measurements, provided a test of the

description, at loop level, of EW interaction in terms of an  $SU(2)_L \times U(1)_Y$  gauge field theory with SSB. The remarkable agreement between directly measured and the 'indirectly' extracted values around the time of the discovery, was a test of the SM at loop level.

Once this was achieved, the same information could be used to obtain constraints on the Higgs mass, now looking at quantum corrections to the  $W$  and  $Z$  mass as well as to the  $Z$  couplings, caused by loops containing the Higgs boson. Finally finding a Higgs boson in 2012 with a mass consistent with these constraints was the biggest success of the SM. Knowledge of QCD was essential in making precision predictions for the Higgs signal and hence to this mass determination. Consistency of the values obtained in theoretical fits with each other and with the experimental measurements, leaves us with no doubt about the correctness of the SM. This tests the correctness of quantum corrections to  $M_W$  coming from the loops containing the  $t$  and  $h$ ; hence of the quantum field theoretic description of the EW interactions as a gauge theory.

Alongside this spectacular testimonial of the correctness of the EW part of the SM, is also the equally impressive demonstration of a highly accurate description of all the CP violating phenomena in terms of the flavor mixing in the quark sector. In the three flavor picture the  $3 \times 3$  Kobayashi-Maskawa matrix is unitary. Making detailed fits of theoretical predictions to a large variety of data on meson mixing and decays, to determine the elements of the Kobayashi-Maskawa matrix with high precision, is an involved exercise as it requires a synthesis of a variety of theoretical tools and high precision data. Since for many of these observables their relationship with the parameters of the SM is given by loop computations, this success too provides a test of the SM as a quantum gauge field theory.

### 27.3 Free Parameters in the SM

The SM is a gauge theory where the requirement of local gauge invariance under chiral isospin transformations results in the minimal couplings to the matter fields. The gauge bosons of the theory acquire a mass via the Higgs mechanism which leads to the prediction of a massive scalar boson in the model. The fermions in the model acquire mass via a Yukawa coupling to this Higgs field. It is worth keeping in mind that the process by which the gauge

bosons acquire a mass derives from the very elegant procedure of spontaneous symmetry breaking and the existence of a finite vacuum expectation value for the Higgs field, so we at least think we understand the origins of the gauge boson masses. The fermion masses are introduced in a totally ad hoc fashion into the model.

The correct gauge group to describe nature is not predicted by the model. The simplest choice consistent with existing phenomenology was suggested in 1968 by Weinberg to be  $SU(2)_L \times U(1)$  and reflected the known  $V - A$  nature of the charged weak interactions. This choice is consistent with all experimental data to date but keep in mind that with this choice, the most striking feature of the weak interactions is simply inserted by hand. Once the gauge group is known, there are 19 free parameters in the model that must be determined. These are listed in the Table below.

Number	Symbol	Description	Value
3 lepton masses	$m_e$	Electron mass	511 keV
	$m_\mu$	Muon mass	105.7 MeV
	$m_\tau$	Tau mass	1.78 GeV
6 quark masses	$m_u$	Up quark mass	1.9 MeV
	$m_d$	Down quark mass	4.4 MeV
	$m_s$	Strange quark mass	87 MeV
	$m_c$	Charm quark mass	1.32 GeV
	$m_b$	Bottom quark mass	4.24 GeV
	$m_t$	Top quark mass	173.5 GeV
3 quark mixing angles	$\theta_{12}$	CKM 12-mixing angle	$13.1^0$
	$\theta_{23}$	CKM 23-mixing angle	$2.4^0$
	$\theta_{13}$	CKM 13-mixing angle	$0.2^0$
2 imaginary quark phase	$\delta$	CKM CP violation phase	0.995
	$\theta_{QCD}$	QCD vacuum angle	0
3 gauge couplings	$g_1$ or $g'$	U(1) gauge coupling	0.357
	$g_2$ or $g$	SU(2) gauge coupling	0.652
	$g_3$ or $g_s$	SU(3) gauge coupling	1.221
2 Higgs parameters	$v$	Higgs vacuum expectation value	246 GeV
	$M_H$	Higgs mass	125.09 GeV

For each commuting set of generators of the group, we have an independent coupling, so there are three gauge couplings  $g_1$ ,  $g_2$ , and  $g_3$  ( $g_s$ ) to be determined experimentally. Here we included  $g_3$  because the Yang-Mills La-



grangian can be extended to include an  $SU(3)$  color symmetry to describe QCD.

The experimentally accessible quantities are the coupling constants  $G_F$ ,  $g_s$  and  $\alpha_{\text{em}}$ , the gauge boson masses,  $M_W$  and  $M_Z$ , and Higgs mass  $M_H$ . The model parameters and the experimental measurables are easily related by the following set of equations:

$$\begin{aligned}
 M_W^2 &= \frac{g_2^2 v^2}{4} , & M_Z^2 &= \frac{g_2^2 v^2}{2 \cos^2 \theta_W} , \\
 e &= g_2 \sin \theta_W , & G_F &= \frac{g_2^2}{8M_W^2} = \frac{1}{2v^2} , \\
 \tan \theta_W &= \frac{g_2}{g_1} , & \sin^2 \theta_W &= 1 - \frac{M_W^2}{M_Z^2} , \\
 \alpha_{\text{em}} &= \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} .
 \end{aligned} \tag{27.2}$$

Very often experimental results are characterized in terms of  $\sin^2 \theta_W$ , which determines the mixing between the neutral  $SU(2)$  and  $U(1)$  gauge fields that result in the physical photon and the  $Z$  boson.

When the matter fields of quarks and leptons are introduced the number of free parameters proliferates appallingly. The fermion-gauge couplings are totally determined by  $\alpha_{\text{em}}$ ,  $g_s$ ,  $G_F$  and  $M_Z$ ; however, the fermion masses coming from the Yukawa coupling of the fermions to the Higgs are all free parameters.

We have another set of parameters to introduce in the form of a rotation matrix. It appears that quark flavor eigenstates of strong interactions are not eigenstates of the weak interactions and we need to experimentally determine the  $3 \times 3$  mixing matrix that rotates one basis into the other. This rotation matrix is called the Kobayashi-Maskawa matrix and it is thought to contain the origins of CP violation. Finally, if neutrinos have mass (and we have no good reason to think they don't) we have to be prepared for the neutrinos to mix as well and there is an equivalent  $3 \times 3$  Kobayashi-Maskawa matrix for lepton sector.

In order for the SM to be completely defined, all these parameters must be measured. Once the model is defined, we can test it and in fact a major part of every high energy physics experiment now has involved testing the predictive power of the SM.

At the end let us note that, in general, one needs to add to the free parameters in the Table the masses and mixing angles of the three known neutrinos, which are just as arbitrary as the others within the confines of the SM. This adds 9 parameters to the 19 listed (3 neutrino masses, 3 neutrino mixing angles and 3 neutrino phases).

**Exercise 27.1:** The leptonic decays  $\mu^+ \rightarrow e^+ \bar{\nu} \nu$  and  $\tau^+ \rightarrow e^+ \bar{\nu} \nu$  both are proceed via the same interaction. If the  $\mu^+$  mean life is  $2.2 \times 10^{-6}$  s, estimate the  $\tau^+$  mean life if the experimental branching ratio for  $\tau$  decay is 16%.

**Exercise 27.2:** Find the mean distance (in the Lab frame) the  $\tau^+$  will travel before decay if it was produced at 29 GeV in the collider process:  $e^+ e^- \rightarrow \tau^+ \tau^-$ .

**Exercise 27.3:** How many visible photons ( $\sim 5,000 \text{ \AA}$ ) does a 100-Watt bulb with 3% efficiency emit per second?

**Exercise 27.4:** Derive the relationship among the particle velocity  $v$ , the index of refraction  $n$  of the medium, and the angle  $\theta$  at which the Cherenkov radiation is emitted relative to the line of flight of the particle.

**Exercise 27.5:** Hydrogen gas at one atmosphere and  $20^\circ \text{ C}$  has an index of refraction  $n = 1 + 1.35 \times 10^{-4}$ . What is the minimum kinetic energy in MeV which an electron would need in order to emit Cherenkov radiation in traversing this medium?

**Exercise 27.6:** A Cherenkov-radiation particle detector is made by fitting a long pipe of one atmosphere and  $20^\circ \text{ C}$  hydrogen gas with an optical system capable of detecting the emitted light and of measuring the angle of emission  $\theta$  to an accuracy of  $\delta\theta = 10^{-3}$  radian. A beam of charged particles with momentum 100 GeV are passed through the counter. Since the momentum is known, measurement of the Cherenkov angle miens a measurement of the rest mass  $m_0$ . For a particle with  $m_0$  near 1 GeV, and to first order in small quantities, what is the fractional error (i.e.  $\delta m_0/m_0$ ) in the determination of  $m_0$  with the Cherenkov counter?

**Exercise 27.7:** Show that the velocity of the  $Z^0$  in the 1-particle creation processes on  $e^+ e^-$  colliders is much larger than  $W^\pm$  bosons.

**Exercise 27.8:** Calculate what the sun's spectrum would be if the mass of the  $W$ -boson were decreased by a factor of two and what it would be if the mass of the  $W$ -boson were increased by a factor of two.

## Part X

### Lecture – Introduction to QCD



## Chapter 28

# History and Outlook

The strong force holds most ordinary matter together because it confines quarks into hadrons such as the proton and neutron. In addition, the strong force binds neutrons and protons to create atomic nuclei. Most of the mass of a common proton or neutron is the result of the strong force field energy; the individual quarks provide only about 1% of the mass of a proton.

The strong interaction is observable at two ranges: on a larger scale, about  $1 - 3 \text{ fm}$  ( $1 \text{ fm} = 10^{-15} \text{ m}$ ), it is the force that binds nucleons together to form the nucleus of an atom. On the smaller scale, less than about  $0.8 \text{ fm}$  (the radius of a nucleon), it is the force (carried by gluons) that holds quarks together to form protons, neutrons, and other hadron particles. In the latter context, it is often known as the color force.

The strong force inherently has such a high strength that hadrons bound by the strong force can produce new massive particles. Thus, if hadrons are struck by high-energy particles, they give rise to new hadrons instead of emitting freely moving radiation. This property of the strong force is called *color confinement*, and it prevents the free "emission" of the strong force: instead, in practice, jets of massive particles are produced.

Strong interactions is mediated by the exchange of massless particles called *gluons* that act between *quarks*, *antiquarks*, and other gluons. Gluons are thought to interact by way of a type of charge called *color charge*. Color charge is analogous to electromagnetic charge, but it comes in three types ( $\pm$  red,  $\pm$  green,  $\pm$  blue) rather than one, which results in a different type

of force, with different behavior.

Quarks are massive spin-1/2 fermions which carry a color charge whose gauging is the content of QCD. Quarks are represented by Dirac fields in the fundamental representation 3 of the gauge group  $SU(3)$ . They also carry electric charge (either  $-1/3$  or  $+2/3$ ) and participate in weak interactions as part of weak isospin doublets. They carry global quantum numbers including the baryon number, which is  $1/3$  for each quark, hypercharge and one of the flavor quantum numbers. Every quark has its own antiquark. The charge of each antiquark is exactly the opposite of the corresponding quark.

Gluons are spin-1 bosons which also carry color charges, since they lie in the adjoint representation 8 of  $SU(3)$ . They have no electric charge, do not participate in the weak interactions, and have no flavor. They lie in the singlet representation 1 of all these symmetry groups.

The rules of quark-gluon interactions are detailed in *Quantum Chromo Dynamics* (QCD), an important part of the SM of particle physics. QCD, which will be considered in this lecture, is a type of non-Abelian gauge theory with  $SU(3)$  symmetry group.

## 28.1 The Yukawa Model

Let us start a brief historical survey in QCD with the mention of the Rutherford experiment, indicated how compact the atomic nucleus is. After the discovery of the neutron by Chadwick, it became necessary to understand how the nucleus is built out of protons and neutrons. The mechanism introduced by Yukawa, the exchange of a massive boson, turned out to be successful, with the discovery of the pion in 1947.

Among the theoretical activity stimulated by the Yukawa model, two points at least deserve attention.

- The mass of the Yukawa boson was constrained by the ratio of the 3-body to the 2-body binding energy;
- At first sight, three potentials were needed to build the nucleus:  $pp$ ,  $pn$  and  $nn$ , and it was natural to seek for some simplification.

A tempting scenario will be that where solely the proton-neutron interaction exists, but this was clearly contradicted by the data on proton scattering. What eventually prevailed was isospin symmetry. With the proton and the neutron forming a doublet, there are only two potentials in the limit where isospin is conserved, one for total isospin  $I = 0$ , and another one for  $I = 1$ .

In 1947 it was discovered the pion and was seen in three charge states ( $\pi^+$ ,  $\pi^0$  and  $\pi^-$ ), forming an isospin triplet. Hence in 1947, we had 5 hadrons: 2 nucleons and 3 pions. But already two more were expected, as the existence of antimatter was predicted as a consequence of the Dirac equation, and the positron was discovered in cosmic rays by Anderson. The antiproton and the antineutron were anticipated as well. A dedicated accelerator was built at Berkeley and the antiproton was, indeed, discovered in 1955, and the antineutron shortly after.

With 7 hadrons, 2 nucleons, 2 antinucleons, and 3 pions, the world of hadronic physics was reasonably sized, and it was necessary to work on the interaction among these few hadrons. However, several complications occurred almost simultaneously.

- First, the Yukawa picture of nuclear forces, though very efficient for the long-range part, faced difficulties at shorter distances. More attraction was needed, and also some spin-orbit component, that neither pion-exchange nor iterated pion-exchange were able to provide. An explicit scalar and vector exchange were needed. While the former was about isospin independent, and thus provided by the exchange of an isoscalar scalar meson, the latter was sought to be different in  $np$  vs.  $pp$  scattering, and thus called for both an isoscalar and an isovector vector meson. The  $\omega$  and  $\rho$  were thus predicted!
- Second, the interaction of pions with nucleons was shown to produce new particles, nucleon resonances, in particular the  $\Delta(1232)$ , which has isospin  $3/2$ , i.e. exists in four possible electric charges. Similarly, proton-nucleon or proton-nucleus scattering, or proton-antiproton annihilation were able to produce several new mesons, which were desired to improve the theory of nuclear forces. These hadrons are not stable, with for instance  $\Delta \rightarrow N + \pi$ , or  $\rho \rightarrow \pi + \pi$ , but were named hadrons as well, baryons or mesons. The  $\Delta$ , at first sight, appears as a consequence of the interaction between  $\pi$  and  $N$ , and the  $\rho$  as a resonance of the  $\pi\pi$  interaction.

Chew and collaborators generalised the scenario, and suggested the concept of *nuclear democracy*: everything is made of everything and any hadron is both a building block and the result of the interaction of the other hadrons. For instance,  $\Delta$  is made of  $N\pi$ ,  $N\pi\pi$ , etc., and, as well  $N$  is made of  $\Delta\pi$ , etc. This gives an infinite set of coupled equations, of which it was hoped one could extract a finite set as a first tractable approximation to the spectrum and to the dynamics. The success was, however, extremely limited. Ball, Scotti and Wong, for instance, stressed that describing mesons as resulting of the nucleon–antinucleon interaction hardly gives the observed "exchange degeneracy" (named after the phenomenology of high-energy scattering), the property that an isoscalar and an isovector mesons with the same quantum numbers have very often the same mass. In the baryon sector, the next state after the  $\Delta$  with isospin  $I = 3/2$  and spin  $J = 3/2$  was sometimes predicted to have  $I = 5/2$  and  $J = 5/2$ !

- Third came strangeness. New particles were observed in the 1950s and 1960s, decaying weakly though massive enough to decay to existing particles, and produced by pairs with strict rules:  $K^+$  together with  $\Lambda$  for instance, but never  $K^-$  together  $\Lambda$ . A new quantum number, strangeness  $S$ , was introduced to summarize the properties of these new particles: strangeness is conserved in production processes by strong interaction, and thus  $\Lambda$  ( $S = -1$ ) can be produced in association with  $K^+$  ( $S = 1$ ), but not  $K^-$  which has  $S = -1$ . On the other hand, strangeness is not conserved in the decay by weak interaction, as  $\Lambda \rightarrow N + \pi$ , or  $\Lambda \rightarrow p + e^- + \bar{\nu}$ . The weak interaction of strange particle was beautifully linked to that involved in ordinary beta decay.

## 28.2 The Sakata Model

It was observed that strange particles do not differ much from the non-strange ones. For instance, the mass of the  $\Lambda$  baryon is about 1.1 GeV, just slightly above that of the nucleon at 0.94 GeV, and the mass of the  $K^*$  (0.89 GeV), is close to that of the vector mesons  $\rho$  and  $\omega$  (about 0.78 GeV). It was thus tempting to put strange and non-strange particles in multiplets generalizing isospin. Since isospin is built on the  $SU(2)$  group, the minimal extension is  $SU(3)$ . Later, it was renamed  $SU(3)_F$ , to differentiate it from the  $SU(3)$  group associated with color. However, there were the exceptions

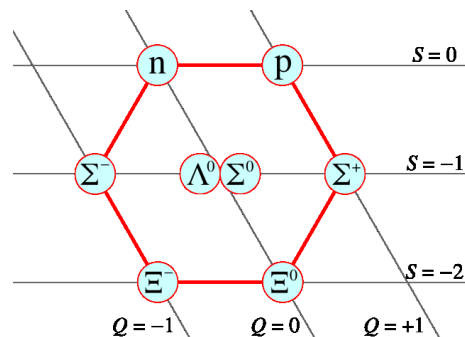


of the  $\pi$  meson, with a mass of about 0.14 GeV (significantly lighter than the mass of the  $K$ , about 0.49 GeV) and the light scalar mesons with long-standing questions about their structure.

A specific model was proposed by Sakata, with  $(n, p, \Lambda)$  as the building blocks of matter, and mesons as baryon-antibaryon pairs. We already mentioned that this picture of mesons is difficult to accommodate with the long-range baryon-baryon interaction as given in the Yukawa picture. But the Sakata model had also problems with baryons. There are too many baryons with low mass. Take for instance the lowest baryons with spin  $J = 1/2$ . Besides  $p$ ,  $n$  and  $\Lambda$ , there is a triplet of singly-strange ( $S = -1$ ) baryons ( $\Sigma^+$ ,  $\Sigma^0$  and  $\Sigma^-$ ) of mass about 1.3 GeV, and a pair of doubly strange ( $S = -2$ ) baryons ( $\Xi^0$  and  $\Xi^-$ ) with mass about 1.5 GeV. In the Sakata model, they should belong to higher representation, and, in a dynamical picture, contain an additional baryon-antibaryon pair. This was hard to believe.

### 28.3 The Eightfold Way

To take care of the problems in Sakata model, Ne'emann and Gell-Mann suggested to keep the  $SU(3)$  group as the basic symmetry, but to put the known  $J = 1/2$  baryons in an octet representation.



The group  $SU(3)$  has eight generators, instead of three for  $SU(2)$  ( $I_{\pm}$  and  $I_3$ ). Each multiplet can be characterized by the dimension of the representation and two generators that commute, which are usually taken as  $I_3$  and strangeness  $S$ , or equivalently the *hypercharge* defined as  $Y = B + S$ , where  $B$  is the baryon number.

$SU(3)$  is rather good symmetry. Its breaking can be described by simple terms, treated at first order, which are proportional to some generators of the group. More elaborate mechanisms were proposed for breaking  $SU(3)$ . One example is the famous Gell-Mann-Okubo formula for the baryon masses

$$M = M_0 + aY + B [I(I + 1) - Y^2/4] , \quad (28.1)$$

where  $M_0$ ,  $a$  and  $B$  are empirical constants related to a given multiplet. From (28.1) one can get the relation

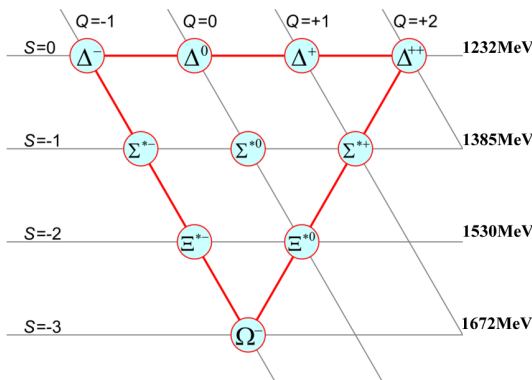
$$2(M_N + M_{\Xi}) = 3M_{\Lambda} + M_{\Sigma} , \quad (28.2)$$

where  $M_N$  stands for nucleon mass, which is in surprisingly good agreement with experiment.

When the eightfold way was proposed, only 9 baryons with spin  $J = 3/2$  were known, with a low mass: four  $\Delta$  of charge ranging from  $-1$  to  $+2$ , three  $\Sigma^*$  with strangeness  $S = -1$  and two  $\Xi^*$  with strangeness  $S = -2$ . At the 1962 Rochester Conference held in Geneva, Gell-Mann pointed out they would fit very well a decuplet representation of  $SU(3)$ , provided the last member does also exist. He named it  $\Omega^-$ , as a kind of ultimate achievement, in the biblical sense. The masses of the 9 existing members being observed to grow linearly with strangeness, i.e. following the pattern

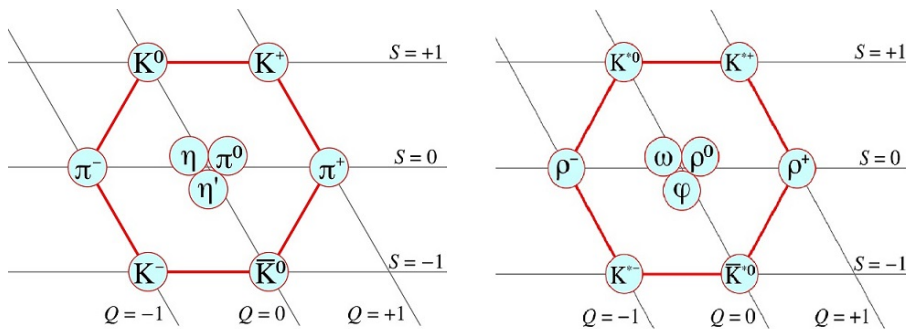
$$M_{\Omega^-} - M_{\Xi^*} = M_{\Xi^*} - M_{\Sigma^*} = M_{\Sigma^*} - M_{\Delta} , \quad (28.3)$$

known as the "equal spacing rule of the decuplet". By including  $M_{\Omega^-} - M_{\Xi^*}$  in this equality, it was predicted the mass of the  $\Omega^-$  at about 1.67 GeV.



Some physicists were skeptical about the possibility of producing and detecting easily the  $\Omega^-$ . It was nevertheless seen in an experiment at Brookhaven led by Samios, with exactly the mass predicted by Gell-Mann. This was at the end of 1963, and published in 1964.

The Eightfold Way organizes into an octet the mesons, as well. The pseudoscalar mesons ( $\pi$ ,  $\eta$ ,  $K$ ,  $\bar{K}$ ) were accommodated in an octet and a singlet. As there is presumably mixing between the singlet and the isoscalar member of the octet, it became customary to talk about "nonet". The same holds for the vector mesons. On the figure below shown the octet and singlet of pseudoscalar (left) and vector (right) mesons.



### 28.3.1 The Fundamental Representation: Quarks

The *Eightfold Way* thus introduced the hadronic octets, singlets and decuplets, but not triplet, which corresponds to the fundamental representation of  $SU(3)$  group. Gell-Mann proposed that the fundamental representation is populated by three yet not discovered (or hypothetical) particles. He of course was fully aware that any representation can be built by combining the fundamental representation,  $\mathbf{3}$ , and its conjugate,  $\bar{\mathbf{3}}$ , in the same way that any value of the spin can be built by adding elementary spins  $1/2$ . The basic reduction of products of the  $\mathbf{3}$  and  $\bar{\mathbf{3}}$  representations that are important for the quark model are:

$$\mathbf{3} \times \mathbf{3} \times \mathbf{3} = 1 + 8 + 8 + 10, \quad \mathbf{3} \times \bar{\mathbf{3}} = 1 + 8. \quad (28.4)$$

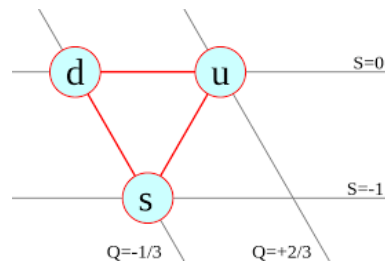
The word *quark* was taken from the sentence *Three Quarks for Muster Mark* in Joyce's "Finnegans Wake". In another context (see next section), the

word *ace* was suggested, but did not prevail. The individual quarks were sometimes named  $p$ ,  $n$  and  $\lambda$ , in reference to the Sakata model, but quickly the naming scheme  $u$ ,  $d$ ,  $s$  was adopted by the community.

The properties of the three quarks are summarized in the table:

$q$	$b$	$I$	$I_3$	$Y$	$S$	$Q$
$u$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{2}{3}$
$d$	$\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	0	$-\frac{1}{3}$
$s$	$\frac{1}{3}$	0	0	$-\frac{2}{3}$	-1	$\frac{2}{3}$

and in the diagram below.

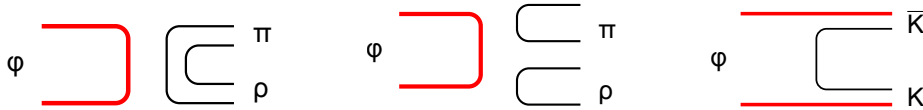


Historians could debate endlessly whether the pioneers considered quarks as a handy mathematical tool to build the representation of  $SU(3)$ , or had already in mind a physical interpretation of the quarks as the constituents of the hadrons. Anyhow, this was one of the major breakthroughs in physics.

## 28.4 The OZI Rule

Another approach was followed by Zweig. In 1962 it was discovered the  $\phi$  meson, of mass about 1.02 GeV. It is an isoscalar, vector meson, like the  $\omega$ , but with peculiar decay patterns. While it could easily decay into pions, it prefers the  $K\bar{K}$  channels, for which the phase-space is meagre (the  $K$  has a mass of about 0.49 GeV). Zweig's explanation is that this favoured decay is dictated by  $\phi$  meson content. He named the constituent *aces*, but we

will call them quarks to conform to the current usage. The idea is that the decay preferentially keeps the existing content. In the modern language, the  $\phi$  decay is described by the diagrams below, Zweig forbidden (left, centre) or allowed (right).

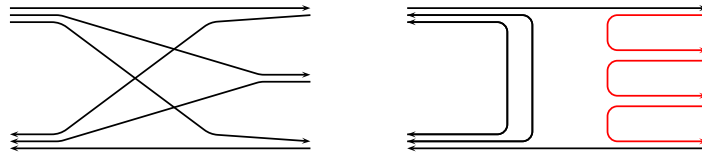


Zweig thereby invented the "Zweig rule", also called Okubo-Iizuka-Zweig (OZI), or (A-Z) rule since many authors contributed, from Alexander to Zweig. As we shall see shortly, a nice surprise was that the rule works even better for heavier quarks.

The rule governing the  $\phi$  decay was extended to other decays and to reaction mechanisms. Quark lines should not start from and end in the same hadron, i.e. disconnected diagrams are suppressed. Instead quark lines should better link two hadrons in the initial or final state.

The OZI rule then got variants. It is sometimes argued that the dominant processes correspond to planar diagrams, while non-planar (but still connected) diagrams are suppressed. For instance, there are dozens of measured branching ratios for antinucleon-nucleon annihilation at rest or in flight. In the diagrams above, the rearrangement diagram (left) clearly keeps the initial constituents, while the diagram on the right is more planar. Clearly, the former does not produce enough kaons, while the second predicts much more kaons than observed.

Note that antiproton annihilation is an indirect evidence for quarks, but this was understood much after the first measurements. If guidance is sought from QED, annihilation has a rather small cross section, because it is a very short-range process. The results obtained at Berkeley for the elastic and annihilation cross-sections of antiproton scattering on nucleons indicated that the latter is larger. To reproduce these results with an empirical (complex) potential, one needs a large size for the annihilation part, about 0.8 fm or more. This was a puzzle. One now understands that the proton and the antiproton are composite, with a size of the order of 0.5 – 1 fm. Hence, when they overlap, they can rearrange their constituents into quark-antiquark pairs. This is similar to the rearrangement collisions in molecular physics, but has little to do with  $e^+e^-$  annihilation in QED.



The hierarchy of the possible diagrams (displayed above) contributing to baryon-antibaryon annihilation requires an extension of the Zweig rule that remains a little controversial.

## 28.5 Quark Approaches

To understand the structure of baryons made of quarks, in 1965 Greenberg suggested a kind of harmonic oscillator as a first approximation to describe the quark motion. He addressed the problem of the statistics and suggested a para-statistics that eventually became the color degree of freedom.

The work of Dalitz was done almost simultaneously. He constructed his first version of the harmonic oscillator model of baryons. As Greenberg, he faced the problem of the statistics of the quark.

Dalitz's work was the starting point of a series of studies about baryons in the harmonic oscillator model with contributions by Horgan, Hey, Kelly, Reinders, Gromes, Stamatescu, Stancu, Cutosky, Isgur and Karl, etc. Potential models not based on harmonic confinement were proposed somewhat later. Note also the contribution by Becchi and Morpurgo about the possibility of describing hadrons made of quarks in a non-relativistic approximation.

### 28.5.1 Heavy Quarks

The physics of kaons has been very stimulating along the years: strangeness led to the quarks, the  $\theta - \tau$  puzzle led to parity violation and to  $K^0 \bar{K}^0$  mixing, whose detailed scrutiny revealed  $CP$  violation. Another problem in the weak decay of kaons inspired Glashow, Iliopoulos and Maiani, who predicted a fourth quark, named *charmed quark* and abbreviated as  $c$ , whose mass should not be too high. In some processes, diagrams involving  $u$  and diagrams involving  $c$  quarks cancel out. This is the GIM mechanism.

A new spectroscopy was thus predicted, with charmed mesons such as  $(c\bar{u})$  and  $(c\bar{c})$ , or charmed baryons such as  $(csu)$ , double-charm baryons, etc. However, when in November 1974 (this was sometimes called the October revolution), the  $J/\psi$  was discovered simultaneously at SLAC and Brookhaven, and the  $\psi'$  shortly after at SLAC, they were not immediately recognized as  $(c\bar{c})$ , i.e. charmonium. The surprise was that they are extremely narrow resonances. We now understand that the Zweig rule works better when the quark become heavier. The spectrum of charmonium was completed with various  $P$  state, the  $\psi'' = \psi(3770)$  which is a  $D$  state, and after some time, with the spin-singlet states. Charmed mesons and baryons were discovered as well, and this sector is now rather rich.

Explicit quark models were developed to describe the  $(c\bar{c})$  spectrum. The charmonium gave a decisive impulse to the quark model in the meson sector. Thanks to Regge and others, we had already an idea about sequences of mesons with increasing spin  $J$ . In the quark model, this corresponds to orbital excitations of the quark-antiquark motion. With charmonium, the new feature is that: (i) The levels are better seen, since the lowest states are narrow; (ii) There is a clear evidence for the radial degree of freedom.

At the time of the discovery of charm, in the 1970s, there were already speculations about a symmetry between quarks and leptons. The light quarks ( $u, d$ ) are the partners of  $(e^-, \nu_e)$  and  $(c, s)$  quarks belong in the family of  $(\mu^-, \nu_\mu)$ . Note that the leptons are ahead. The  $\mu^-$  was discovered in 1936, and was completely unexpected. The  $\tau$  lepton was discovered at SLAC in 1977. The partners of the  $(\tau^-, \nu_\tau)$  pairs were thus anticipated and named  $(t, b)$ , as *top* and *bottom*. And a spectroscopy of hadrons containing  $b$  or  $t$ , or both, was predicted. However, at a Conference in Hamburg, a German minister who had some knowledge of English, suggested to replace *bottom* by *beauty*. And *top* was renamed *truth*, but this is not very often used.

In 1977, Lederman, who missed the charmonium by a small margin, did an experiment similar to Ting's, but with a more powerful beam and an improved detector. In 1977 he announced the discovery of the  $\Upsilon$  and  $\Upsilon'$  particles, immediately interpreted as  $(b\bar{b})$  bound states. Lederman noticed that, within the precision of his measurement for the masses,

$$\Upsilon' - \Upsilon \simeq \psi' - J/\psi, \quad (28.5)$$

and submitted to local theorists the question whether there exists a potential such that changing the reduced mass keep the spacings  $\Delta E$  unchanged

(remember that  $\Delta E \propto m$  for the Coulomb potential and  $\Delta E \propto 1/\sqrt{m}$  for the harmonic oscillator). Quigg and Rosner found that the logarithmic potential has this property for all spacings. In fact, the logarithmic potential was already used in empirical speculations about the bottomonium  $b\bar{b}$ , but this property was not stressed clearly enough.

### 28.5.2 A First Hint of Colour

The story of the color degree of freedom arguably can be started with the discovery of  $\Delta^{++}$  baryon in 1951. The eightfold way refers to the classification of the lowest-lying pseudoscalar mesons and spin-1/2 baryons within octets in  $SU(3)$ -flavour space ( $u, d, s$ ). The  $\Delta^{++}$  is part of a spin-3/2 baryon decuplet, a *tenfold way* in this terminology.

In the context of the quark model - which first had to be developed, successively joining together the notions of spin, isospin, strangeness, and the eightfold way - the flavor and spin content of the  $\Delta^{++}$  baryon is:

$$|\Delta^{++}\rangle = |u_{\uparrow} u_{\uparrow} u_{\uparrow}\rangle , \quad (28.6)$$

clearly a highly symmetric configuration. However, since the  $\Delta^{++}$  is a fermion, it must have an overall antisymmetric wave function.

In the beginning of 1965, Bogolyubov, Struminsky and Tavkhelidze wrote a preprint (with a discussion of the earlier idea of Struminsky) on the necessity of additional quark quantum degree of freedom in connection with  $\Omega^{-}$  hyperon. The situation with  $\Omega^{-}$  (composed of three  $s$  quarks with parallel spins) also was peculiar, as for the case of  $\Delta^{++}$  - without additional degree of freedom such combination is forbidden by the Pauli exclusion principle.

In 1964-65 Greenberg and Han-Nambu independently resolved the problem of  $\Delta^{++}$  by proposing that quarks possess an additional  $SU(3)$  gauge degree of freedom, later called color charge. The  $\Delta^{++}$  wave function can now be made antisymmetric by arranging its three quarks antisymmetrically in this new degree of freedom,

$$|\Delta^{++}\rangle = \epsilon^{ijk} |u_{i\uparrow} u_{j\uparrow} u_{k\uparrow}\rangle , \quad (28.7)$$

hence solving the mystery. Han and Nambu noted that quarks might interact via an octet of vector gauge bosons: the gluons.



More direct experimental tests of the number of colors,  $N_C$ , were provided first by measurements of the decay width of  $\pi^0 \rightarrow \gamma\gamma$  decays, which is proportional to  $N_C^2$ , and later by the famous  $R$ -ratio in  $e^+e^-$  collisions

$$R = \frac{\sigma(e^+e^- \rightarrow q\bar{q})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}, \quad (28.8)$$

which is proportional to  $N_C$ .

## 28.6 Development of QCD

Since free quark searches consistently failed to turn up any evidence for the new particles, and because an elementary particle back then was defined as a particle which could be separated and isolated, Gell-Mann often said that quarks were merely convenient mathematical constructs, not real particles. The meaning of this statement was usually clear in context: He meant quarks are confined, but he also was implying that the strong interactions could probably not be fully described by QFT.

Feynman argued that experiments showed quarks are real particles: he called them partons (since they were parts of hadrons). By particles, Feynman meant objects which travel along paths, elementary particles in a field theory.

The difference between Feynman's and Gell-Mann's approaches reflected a deep split in the theoretical physics community. Feynman thought the quarks have a distribution of position or momentum, like any other particle, and he (correctly) believed that the diffusion of parton momentum explained diffractive scattering. Although Gell-Mann believed that certain quark charges could be localized, he was open to the possibility that the quarks themselves could not be localized because space and time break down. This was the more radical approach of  $S$ -matrix theory.

Bjorken proposed that point-like partons would imply certain relations in deep inelastic scattering of electrons and protons, which were verified in experiments at SLAC in 1969. This led physicists to abandon the  $S$ -matrix approach for the strong interactions.

In 1973 the concept of color as the source of a "strong field" was developed into the theory of QCD by Fritzsche and Leutwyler, together with Gell-Mann. In particular, they employed the general field theory developed in 1954 by

Yang and Mills, in which the carrier particles of a force can themselves radiate further carrier particles (this is different from QED, where the photons that carry the electromagnetic force do not radiate further photons).

The discovery of asymptotic freedom in the strong interactions by Gross, Politzer and Wilczek allowed physicists to make precise predictions of the results of many high energy experiments using the QFT technique of perturbation theory. Evidence of gluons was discovered in three-jet events at PETRA in 1979. These experiments became more and more precise, culminating in the verification of perturbative QCD at the level of a few percent at the LEP in CERN.

The other side of asymptotic freedom is confinement. Since the force between color charges does not decrease with distance, it is believed that quarks and gluons can never be liberated from hadrons. This aspect of the theory is verified within lattice QCD computations, but is not mathematically proven. One of the Millennium Prize Problems announced by the Clay Mathematics Institute requires a claimant to produce such a proof. Other aspects of non-perturbative QCD are the exploration of phases of quark matter, including the quark-gluon plasma.

The relation between the short-distance particle limit and the confining long-distance limit is one of the topics recently explored using string theory, the modern form of  $S$ -matrix theory.

## 28.7 Distinguished Properties

As it was already mentioned QCD exhibits two peculiar properties: asymptotic freedom and color confinement. Let's briefly describe these features.

### 28.7.1 Color Confinement

*Color Confinement* is a consequence of the constant force between two color charges as they are separated: In order to increase the separation between two quarks within a hadron, ever-increasing amounts of energy are required. Eventually this energy produces a quark-antiquark pair, turning the initial hadron into a pair of hadrons instead of producing an isolated color charge.

Although analytically unproven, color confinement is well established from lattice QCD calculations and decades of experiments.

Detailed computations with the QCD Lagrangian show that the effective potential between a quark and its anti-quark in a meson contains a term that increases in proportion to the distance between the quark and anti-quark ( $\propto r$ ), which represents some kind of "stiffness" of the interaction between the particle and its anti-particle at large distances, similar to the entropic elasticity of a rubber band. This leads to confinement of the quarks to the interior of hadrons, i.e. mesons and nucleons, with typical radii corresponding to former "bag models" of the hadrons. The order of magnitude of the "bag radius" is  $\sim 1$  fm. Moreover, the above-mentioned stiffness is quantitatively related to the so-called "area law" behavior of the expectation value of the Wilson loop, which is proportional to the area enclosed by the loop.

Note that Wilson loop is an important theoretical concept in QCD. In lattice approach, the final term of the QCD Lagrangian is discretized via Wilson loops, and more generally the behavior of Wilson loops can distinguish confined and deconfined phases.

### 28.7.2 Asymptotic Freedom and Duality

QCD has the property of *asymptotic freedom*, i.e. the running coupling becomes weak at high energies, or short distances. At low energies, or large distances, QCD becomes strongly coupled (infrared slavery), presumably leading to the confinement of quarks and gluons. The asymptotic freedom of QCD was discovered in 1973 by David Gross and Frank Wilczek, and independently by David Politzer in the same year. For this work all three shared the 2004 Nobel Prize in Physics.

Asymptotic freedom means that at large energy, or short distances, there is practically no interaction between quarks and gluons, as the energy scale of those interactions increases (and the corresponding length scale decreases). This is in contrast (one would say dual) to what one is used to, since usually one connects the absence of interactions with large distances. As mentioned by Wegner, a solid state theorist, who introduced 1971 simple gauge invariant lattice models, the high-temperature behavior of the original model, e.g. the strong decay of correlations at large distances, corresponds to the low-temperature behavior of the dual model, namely the asymptotic decay

of non-trivial correlations, e.g. short-range deviations from almost perfect arrangements, for short distances.

## 28.8 The Strong Coupling

To first approximation, QCD is scale invariant, if one 'zooms in' on a QCD jet, one will find a repeated self-similar pattern of jets within jets, reminiscent of fractals. In the context of QCD, this property was originally called light-cone scaling, or Bjørken scaling. This type of scaling is closely related to the class of angle-preserving symmetries, called conformal symmetries. In physics today, the terms *conformal* and *scale invariant* are used interchangeably. Strictly speaking, conformal symmetry is more restrictive than just scale invariance, but examples of scale-invariant field theories that are not conformal are rare. Conformal invariance is a mathematical property of several *QCD-like* theories which are now being studied (such as  $N = 4$  SUSY relatives of QCD). It is also related to the physics of so-called *unparticles*.

If the strong coupling did not run, Bjørken scaling would be absolutely true and QCD would be a theory with a fixed coupling, the same at all scales. This simplified picture already captures some of the most important properties of QCD. In the limit of exact Bjørken scaling (QCD at fixed coupling) properties of high-energy interactions are determined only by dimensionless kinematic quantities, such as scattering angles (pseudo-rapidities) and ratios of energy scales. Originally, the observed approximate agreement with this was used as a powerful argument for point-like substructure in hadrons; since measurements at different energies are sensitive to different resolution scales, independence of the absolute energy scale is indicative of the absence of other fundamental scales in the problem and hence of point-like constituents. For applications of QCD to high-energy collider physics, an important consequence of Bjørken scaling is thus that the rate of *bremstrahlung jets*, with a given transverse momentum, scales in direct proportion to the hardness of the fundamental partonic scattering process they are produced in association with. This agrees well with our intuition about accelerated charges; the harder you 'kick' them, the harder the radiation they produce.

For instance, in the limit of exact scaling, a measurement of the rate of 10-GeV jets produced in association with an ordinary  $Z$  boson could be used as a direct prediction of the rate of 100-GeV jets that would be produced in

association with a 900-GeV  $Z'$  boson, and so forth. Our intuition about how many bremsstrahlung jets a given type of process is likely to have should therefore be governed first and foremost by the ratios of scales that appear in that particular process, as has been highlighted in a number of studies focusing on the mass and  $p_{\perp}$  scales appearing, e.g. beyond the SM physics processes Björken scaling is also fundamental to the understanding of jet substructure in QCD.

On top of the underlying scaling behavior, the running coupling will introduce a dependence on the absolute scale, implying more radiation at low scales than at high ones. The running is logarithmic with energy, leads to the asymptotic freedom. More correctly, it is the potential which grows linearly, while the force becomes constant.

Among the consequences of asymptotic freedom is that perturbation theory becomes better behaved at higher absolute energies, due to the effectively decreasing coupling. Perturbative calculations for our 900-GeV  $Z'$  boson from before should therefore be slightly faster converging than equivalent calculations for the 90-GeV one. Furthermore, since the running of  $\alpha_s$  explicitly breaks Björken scaling, we also expect to see small changes in jet shapes and in jet production ratios as we vary the energy. For instance, since high- $p_{\perp}$  jets start out with a smaller effective coupling, their intrinsic shape (irrespective of boost effects) is somewhat narrower than for low- $p_{\perp}$  jets, an issue which can be important for jet calibration.

As a final remark on asymptotic freedom, note that the decreasing value of the strong coupling with energy must eventually cause it to become comparable to the EW one at some energy scale. Beyond that point, which may lie at energies of order  $10^{15} - 10^{17}$  GeV (though it may be lower if as yet undiscovered particles generate large corrections to the running), we do not know what the further evolution of the combined theory will actually look like, or whether it will continue to exhibit asymptotic freedom.

It is sometimes stated that QCD only has a single free parameter, the strong coupling. However, even in the perturbative region, the beta function depends explicitly on the number of quark flavors, as we have seen, and thereby also on the quark masses. Furthermore, in the non-perturbative region the value of the perturbative coupling gives little or no insight into the behavior of the full theory. Instead, universal functions (such as parton densities, form factors, fragmentation functions, etc.), effective theories (such as the *Operator Product Expansion*, *Chiral Perturbation Theory*, or *Heavy Quark*

*Effective Theory*), or phenomenological models (such as *Regge Theory* or the *String and Cluster Hadronisation Models*) must be used, which in turn depend on additional non-perturbative parameters.

For some of these questions, such as hadron masses, lattice QCD can furnish important additional insight, but for multi-scale and/or time-evolution problems, the applicability of lattice methods is still severely restricted; the lattice formulation of QCD requires a Wick rotation to Euclidean space. The time-coordinate can then be treated on an equal footing with the other dimensions, but intrinsically Minkowskian problems, such as the time evolution of a system, are inaccessible. The limited size of current lattices also severely constrain the scale hierarchies that it is possible to 'fit' between the lattice spacing and the lattice size.

## 28.9 Methods of QCD

When probed at very short wavelengths, QCD is essentially a theory of free 'partons' (quarks and gluons), which only scatter off one another through relatively small quantum corrections, that can be systematically calculated. But at longer wavelengths, of order the size of the proton  $\sim 1 \text{ fm} = 10^{-15} \text{ m}$ , we see strongly bound towers of hadron resonances emerge, with string-like potentials building up if we try to separate their partonic constituents. Due to our inability to perform analytic calculations in strongly coupled field theories, QCD is therefore still only partially solved.

The consequence for collider physics is that some parts of QCD can be calculated in terms of the fundamental parameters of the Lagrangian, whereas others must be expressed through models or functions whose effective parameters are not a priori calculable but which can be constrained by fits to data. However, even in the absence of a perturbative expansion, there are still several strong theorems which hold, and which can be used to give relations between seemingly different processes. This is the reason it makes sense to measure the partonic substructure of the proton in  $ep$  collisions and then re-use the same parametrizations for  $pp$  collisions. Thus the loss of a factorized perturbative expansion is not equivalent to a total loss of predictivity.

An alternative approach would be to give up on calculating QCD and use leptons instead. Formally, this amounts to summing inclusively over strong-

interaction phenomena, when such are present. While such a strategy might succeed in replacing what we do know about QCD by 'unity'.

Furthermore, QCD is the richest gauge theory we have so far. Its emergent phenomena, unitarity properties, color structure, non-perturbative dynamics, quantum vs. classical limits, interplay between scale-invariant and scale-dependent properties, and its wide range of phenomenological applications, are still very much topics of active investigation.

In addition, or perhaps as a consequence, the field of QCD is currently experiencing something of a revolution. On the perturbative side, new methods to compute scattering amplitudes with very high particle multiplicities are being developed, together with advanced techniques for combining such amplitudes with all-orders re-summation frameworks. On the non-perturbative side, the wealth of data on soft-physics processes from the LHC is forcing us to reconsider the reliability of the standard fragmentation models, and heavy-ion collisions are providing new insights into the collective behavior of hadronic matter. The study of cosmic rays impinging on the Earth's atmosphere challenges our ability to extrapolate fragmentation models from collider energy scales to the region of ultra-high energy cosmic rays. And finally, dark matter annihilation processes in space may produce hadrons, whose spectra are sensitive to the modeling of fragmentation.

Detailed analysis of the content of the QCD is complicated and various techniques have been developed. Some of them are discussed briefly below:

- *Perturbative QCD*. This approach is based on asymptotic freedom, which allows perturbation theory to be used accurately in experiments performed at very high energies. Although limited in scope, this approach has resulted in the most precise tests of QCD to date.
- *Lattice QCD*. Among non-perturbative approaches to QCD, the most well established one is lattice QCD. This approach uses a discrete set of spacetime points (called the lattice) to reduce the analytically intractable path integrals of the continuum theory to a very difficult numerical computation which is then carried out on supercomputers. While it is a slow and resource-intensive approach, it has wide applicability, giving insight into parts of the theory inaccessible by other means, in particular into the explicit forces acting between quarks and antiquarks in a meson. However, the numerical sign problem makes it difficult to use lattice methods to study QCD at high density and low

temperature (e.g. nuclear matter or the interior of neutron stars).

- *1/N<sub>c</sub> expansion.* A well-known approximation scheme, the 1/N<sub>c</sub> expansion, starts from the idea that the number of colors N<sub>c</sub> is infinite, and makes a series of corrections to account for the fact that it is not. Until now, it has been the source of qualitative insight rather than a method for quantitative predictions. Modern variants include the AdS/CFT approach.
- *Effective theories.* For specific problems effective theories may be written down which give qualitatively correct results in certain limits. In the best of cases, these may then be obtained as systematic expansions in some parameter of the QCD Lagrangian.

One such effective field theory is *Chiral Perturbation Theory*, which is the QCD effective theory at low energies. More precisely, it is a low energy expansion based on the chiral SSB of QCD, which is an exact symmetry when quark masses are equal to zero, but for the *u*, *d* and *s* quark, which have small mass, it is still a good approximate symmetry. Depending on the number of quarks which are treated as light, one uses either *SU(2)* or *SU(3)* chiral models.

Other effective theories are *Heavy Quark Effective Theory* (which expands around heavy quark mass near infinity), and *Soft-Collinear Effective Theory* (which expands around large ratios of energy scales). In addition to effective theories, models like the Nambu-Jona-Lasinio model and the chiral model are often used when discussing general features.

## 28.10 Confirmation and Outlook

A large body of experimental evidence for QCD has been gathered over the years. In particular, several good quantitative tests of perturbative QCD exist:

- The running of the QCD coupling;
- Scaling violation in polarized/unpolarized deep inelastic scatterings;
- Vector boson productions at colliders (included Drell-Yan process);



- Direct photons productions in hadronic collisions;
- Jet cross sections in colliders;
- Event shape observables;
- Heavy-quark productions in colliders.

Quantitative tests of non-perturbative QCD are fewer, because the predictions are harder to make. The best is probably the running of the QCD coupling as probed through lattice computations of heavy-quarkonium spectra. There is a claim about the mass of the heavy meson  $B_c$ . Other non-perturbative tests are currently at the level of 5% at best. Continuing work on masses and form factors of hadrons and their weak matrix elements are promising candidates for future quantitative tests. The whole subject of quark matter and the quark-gluon plasma is a non-perturbative test bed for QCD which still remains to be properly exploited.

One qualitative prediction of QCD is that there exist composite particles made solely of gluons called glueballs that have not yet been definitively observed experimentally. A definitive observation of a glueball with the properties predicted by QCD would strongly confirm the theory. In principle, if glueballs could be definitively ruled out, this would be a serious experimental blow to QCD. But scientists are unable to confirm or deny the existence of glueballs definitively, despite the fact that particle accelerators have sufficient energy to generate them.

On the way from the early days of the quark model to the recent state of art, there are many beautiful and decisive contributions. Let us list of few items that remains about 50 years after the first speculations on quarks:

- *$SU(3)_F$  symmetry.* The  $SU(3)$  flavor symmetry is now becoming somewhat controversial. Sometimes, it is almost exact, e.g. the case for the decay rates of  $J/\psi$ . In other cases,  $SU(3)_F$  looks badly broken. For instance, if one looks at  $K$  to  $\pi$  production in high-energy collisions, one finds a large "strangeness suppression". Perhaps the elementary couplings for  $u\bar{u}$  or  $s\bar{s}$  production are very similar, but the small differences in masses are amplified by a tunneling factor which is exponential.
- *From flavor symmetry to flavor independence.* In hadron models, it is delicate to attempt a blind extension of  $SU(3)_F$  to  $SU(4)$  or  $SU(5)$  to

include charm and beauty. The breaking is too large to be treated as a perturbation. A new understanding of the symmetry has been provided by the quark model. The basic interaction is flavor independent, or has a well-defined flavor dependence in the spin-spin and spin-orbit terms. Then the same basic interaction is used to build the various  $(q\bar{q}')$  mesons and  $(qq'q'')$  baryons.

- *The bootstrap, or nuclear democracy concept.* This idea was conceptually superb, but operationally somewhat inefficient. Still, a kind of duality exists: for instance, it is natural to consider the  $\Delta$  baryon in some context as the partner of  $N = (qqq)$  nucleon with all quark spins aligned, but to describe the pion-nucleus scattering, there is nothing better than  $\Delta$  as a  $\pi - N$  resonance. When the quark model was invented, it was considered as a relief, with at last a systematic and coherent scheme to describe the hadrons. Nowadays, for any hadron  $a$  of any mass  $m$ , one finds a pair of existing hadrons  $b$  and  $c$  and an effective Lagrangian coupling  $a$ ,  $b$  and  $c$ , such that  $a$  appears as a kind of molecule. The overall picture is seemingly lost.
- *From quark models to QCD.* This will be covered in some of the next chapters. It is fascinating how a somewhat empirical model, but based on the acute observation of facts, became a beautiful theory.
- *A premium to simplicity.* The quark model, even in its simplest variant, works rather well. Presumably, the quark model incorporates a lot of subtle dynamical effects in an effective way, and enables one to make successful predictions.
- *Hadron dynamics beyond hadron decay.* An ambitious application of the quark model deals with the hadron-hadron interaction. This is the continuum part of the sections devoted to loop corrections and to multi-quark spectroscopy.

**Exercise 28.1:** By considering the symmetry of the wave function, explain why the existence of the  $\Omega^-(sss)L = 0$  baryon provides evidence for a degree of freedom in addition to space + spin + flavor.

**Exercise 28.2:** Why can't the "ninth gluon" be the photon?

## Chapter 29

# Structure of QCD

As we mention in previous chapter, necessity of the theory for strong interactions had raised in 1950s, when experimental particle physics discovered a large and ever-growing number of hadrons. It seemed that such a large number of particles could not all be fundamental. At first the particles were classified by charge and isospin by Wigner and Heisenberg; then, in 1953-56, according to strangeness by Gell-Mann and Nishijima. To gain greater insight, the hadrons were sorted into groups having similar properties and masses using the eightfold way, invented in 1961 by Gell-Mann and Ne'eman. Gell-Mann and Zweig, correcting an earlier approach of Sakata, went on to propose in 1963 that the structure of the groups could be explained by the existence of three flavors of smaller particles inside the hadrons: the quarks.

### 29.1 Symmetries of the Model

QCD is a gauge theory of the  $SU(3)$  gauge group obtained by taking the color charge to define a local symmetry. This symmetry acts on the different colors of quarks, and this is an exact gauge symmetry mediated by the gluons.  $SU(3)$  is the Special Unitary group in 3 (complex) dimensions, whose elements are the set of unitary  $3 \times 3$  matrices with determinant one. Since there are 9 linearly independent unitary complex matrices (a complex  $N \times N$  matrix has  $2N^2$  degrees of freedom, on which unitarity provides  $N^2$  constraints), one of which has determinant  $-1$ , there are a total of 8 in-

dependent directions in this matrix space, corresponding to eight different generators as compared with the single one of QED. For  $SU(3)$  the dimension of the adjoint, or vector representation is equal to the number of generators,  $N^2 - 1 = 8$ , while the dimension of the fundamental representation is the degree of the group,  $N = 3$ .

In the context of QCD, we normally represent  $SU(3)$  group using the fundamental representation, in which its generators  $\lambda^a$  ( $a = 1, 2, \dots, 8$ ) appear as a set of eight traceless and hermitian Gell-Mann matrices:

$$\begin{aligned}\lambda^i &= \begin{pmatrix} \sigma^i & 0 \\ 0 & 0 \end{pmatrix}, & \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & (i = 1, 2, 3) \\ \lambda^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},\end{aligned}$$

where  $\sigma^i$  denotes three  $2 \times 2$  Pauli matrices. These matrices can operate both on each other (representing combinations of successive gauge transformations) and on a set of 3-vectors, the latter of which represent quarks in color space; the quarks are triplets under  $SU(3)$ . The matrices can be thought of as representing gluons in color space (or, more precisely, the gauge transformations carried out by gluons), hence there are eight different gluons, which are octets under  $SU(3)$ .

The structure constants of  $SU(3)$  are listed to the right. They define the adjoint representation and are related to the fundamental representation generators via the commutator relations

$$\lambda^a \lambda^b - \lambda^b \lambda^a = i f^{abc} \lambda_c \quad (29.1)$$

( $a, b, c = 1, \dots, 8$ ). Thus one can express color space on a basis of  $\lambda^a$  or via the structure constants.

#### **SU(3) Structure Constants**

$$f_{123} = 1$$

$$f_{147} = f_{246} = f_{257} = f_{345} = \frac{1}{2}$$

$$f_{156} = f_{367} = -\frac{1}{2}$$

$$f_{458} = f_{678} = \frac{\sqrt{3}}{2}$$

Antisymmetric in all indices

All other  $f_{abc} = 0$

QCD exhibits also the second, different  $SU(3)$  symmetry. Since the strong interaction does not discriminate between different flavors of quark, there

is also a flavor symmetry which rotates different flavors of quarks to each other, or  $SU(3)_F$ , which is an approximate symmetry of the vacuum of QCD, and is not a fundamental symmetry at all. This symmetry is an accidental consequence of the small mass of the three lightest quarks and is broken by the differing these masses. The approximate flavor symmetries do have associated gauge bosons, observed particles like  $\rho$  and  $\Omega$ , but these particles are nothing like the gluons and they are not massless. They are emergent gauge bosons in an approximate string description of QCD.

In the QCD vacuum there are vacuum condensates of all the quarks whose mass is less than the QCD scale. This includes the up and down quarks, and to a lesser extent the strange quark, but not any of the others. The vacuum is symmetric under  $SU(2)$  isospin rotations of up and down, and to a lesser extent under rotations of up, down and strange, or full flavor group  $SU(3)_F$ , and the observed particles make isospin and  $SU(3)_F$  multiplets.

Note that there are additional global symmetries whose definitions require the notion of chirality, discrimination between left and right-handed.

### 29.1.1 Color States

The underlying  $SU(3)$  group theory can be used to find out which color states we can obtain by combinations of the red ( $R$ ), green ( $G$ ) and blue ( $B$ ) quarks and gluons. The simplest example of this is the combination of a quark and antiquark. We can form a total of nine different color-anticolor combinations, which fall into two irreducible representations of  $SU(3)$ :

$$3 \times \bar{3} = 8 \oplus 1 . \quad (29.2)$$

The singlet here corresponds to the symmetric wave function

$$\frac{1}{\sqrt{3}} (|R\bar{R}\rangle + |G\bar{G}\rangle + |B\bar{B}\rangle) , \quad (29.3)$$

which is invariant under  $SU(3)$  transformations (the definition of a singlet). The other eight linearly independent combinations (which can be represented by one for each Gell-Mann matrix, with the singlet corresponding to the identity matrix) transform into each other under  $SU(3)$ . Thus, although we sometimes talk about color-singlet states as being made up (e.g. red-antired) that is not quite precise language. The actual state  $|R\bar{R}\rangle$  is not a pure color singlet. Although it does have a non-zero projection onto the singlet wave

function above, it also has non-zero projections onto the two members of the octet that correspond to the diagonal Gell-Mann matrices,  $\lambda_3$  and  $\lambda_8$ . Intuitively, one can also easily realise this by noting that an  $SU(3)$  rotation of  $|R\bar{R}\rangle$  would in general turn it into a different state, say  $|B\bar{B}\rangle$ , whereas a true color singlet would be invariant. Finally, we can also realise from (29.2) that a random (color-uncorrelated) quark-antiquark pair has a  $1/N^2 = 1/9$  chance to be in an overall color-singlet state; otherwise it is in an octet.

Similarly, there are also nine possible quark-quark (or antiquark-antiquark) combinations, six of which are symmetric under interchange of the two quarks and three of which are antisymmetric:

$$6 = \begin{pmatrix} |RR\rangle \\ |GG\rangle \\ |BB\rangle \\ \frac{1}{\sqrt{2}}(|RG\rangle + |GR\rangle) \\ \frac{1}{\sqrt{2}}(|GB\rangle + |BG\rangle) \\ \frac{1}{\sqrt{2}}(|BR\rangle + |RB\rangle) \end{pmatrix}, \quad \bar{3} = \begin{pmatrix} \frac{1}{\sqrt{2}}(|RG\rangle - |GR\rangle) \\ \frac{1}{\sqrt{2}}(|GB\rangle - |BG\rangle) \\ \frac{1}{\sqrt{2}}(|BR\rangle - |RB\rangle) \end{pmatrix}. \quad (29.4)$$

The members of the sextet transform into (linear combinations of) each other under  $SU(3)$  transformations, and similarly for the members of the antitriplet, hence neither of these can be reduced further. The breakdown into irreducible  $SU(3)$  multiplets is therefore

$$3 \times 3 = 6 \oplus \bar{3}. \quad (29.5)$$

Thus, an uncorrelated pair of quarks has a  $1/3$  chance to add to an overall anti-triplet state (corresponding to coherent superpositions like  $R + G = \bar{B}$ ); otherwise it is in an overall sextet state. In the context of hadronisation models, this coherent superposition of two quarks in an overall antitriplet state is sometimes called a *diquark* (at low  $m_{qq}$ ) or a *string junction* (at high  $m_{qq}$ ); it corresponds to the antisymmetric  $R + G = \bar{B}$  combination needed to create a baryon wavefunction.

Note that the emphasis on the quark-(anti)quark pair being uncorrelated is important; production processes that correlate the produced partons, like  $Z \rightarrow q\bar{q}$  or  $g \rightarrow q\bar{q}$ , will project out specific components (here the singlet and octet, respectively). Note also that if the quark and antiquark are on opposite sides of the universe (i.e. living in two different hadrons), the QCD dynamics will not care what overall color state they are in. So for the formation of multi-partonic states in QCD the spatial part of the wave functions (causality

at the very least) will also play a role. Here, we are considering only the color part of the wave functions. Some additional examples are

$$\begin{aligned}
 8 \times 8 &= 27 \oplus 10 \oplus \bar{10} \oplus 8 \oplus 8 \oplus 1, \\
 3 \times 8 &= 15 \oplus 6 \oplus 3, \\
 3 \times 6 &= 10 \oplus 8, \\
 3 \times 3 \times 3 &= (6 \oplus \bar{3}) \times 3 = 10 \oplus 8 \oplus 8 \oplus 1.
 \end{aligned} \tag{29.6}$$

Physically, the 27 in the first line corresponds to a completely incoherent addition of the color charges of two gluons; the decuplets are slightly more coherent (with a lower total color charge), the octets yet more, and the singlet corresponds to the combination of two gluons that have precisely equal and opposite color charges, so that their total charge is zero.

## 29.2 The Lagrangian of QCD

QCD, the fundamental theory of the strong interactions describing a huge variety of phenomena in hadronic and nuclear physics, is an  $SU(3)$  gauge theory described by the Lagrangian density that fits into a single line

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} + \bar{\psi}_q^i (i\gamma^\mu) (D_\mu)_{ij} \psi_q^j - \mathcal{M} \bar{\psi}_q^i \psi_{qi}. \tag{29.7}$$

Here

$$\begin{aligned}
 F_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \quad (a, b, c = 1, \dots, 8) \\
 D_\mu &= \partial_\mu \mathbf{1} - \frac{i}{2} \lambda^a G_\mu^a,
 \end{aligned} \tag{29.8}$$

where  $F_{\mu\nu}^a$  is the field strength tensor for the gluon fields  $G_\mu^a$  with (adjoint) color index  $a$  and  $D_\mu$  is the covariant derivative in QCD. The quantity  $g_s$  denotes the strong coupling (related to  $\alpha_s$  by  $g_s^2 = 4\pi\alpha_s$ ) and  $\lambda^a$  are the  $3 \times 3$  Gell-Mann matrices (29.1), which are the generators of the  $SU(3)$  color group, and are normalized by

$$\text{Tr } \lambda^a \lambda^b = 2\delta^{ab}. \tag{29.9}$$

In (29.7), the Dirac-spinor  $\psi_q^i$  with (fundamental) color index  $i$ ,

$$\psi_q = (\psi_{qR}, \psi_{qG}, \psi_{qB})^T, \tag{29.10}$$

is a vector in the flavor space, with the components describing  $u$ ,  $d$ ,  $s$ ,  $c$ ,  $b$  and  $t$  quarks. Further,  $\mathcal{M}$  denotes the quark mass matrix

$$\mathcal{M} = \text{diag}(m_u, m_d, m_s, m_c, m_b, m_t) . \quad (29.11)$$

There are no bare mass terms for the quarks in (29.7). These would be allowed by QCD alone, but are forbidden by the chiral symmetry of the EW part of the SM. The quark masses are generated by SSB, i.e. by the standard Higgs mechanism or similar.

The Lagrangian of QCD is invariant under local  $SU(3)$  color gauge group transformations

$$\begin{aligned} \psi'(x) &= \Lambda(x)\psi(x) , & \bar{\psi}'(x) &= \bar{\psi}(x)\Lambda(x)^\dagger , \\ G'_\mu(x) &= \Lambda(x)G_\mu(x)\Lambda(x)^\dagger - \partial_\mu\Lambda(x)\Lambda(x)^\dagger , \end{aligned} \quad (29.12)$$

where  $\Lambda(x) \in SU(3)$ . Under these transformations, the stress tensor transforms covariantly

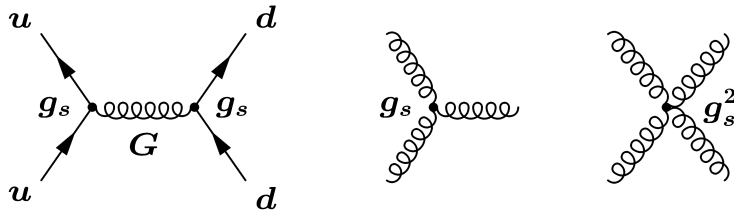
$$F'_{\mu\nu}(x) = \Lambda(x)F_{\mu\nu}(x)\Lambda(x)^\dagger , \quad (29.13)$$

where we defined the matrix-valued fields

$$F_{\mu\nu} = -\frac{i}{2}\lambda^a F_{\mu\nu}^a = \partial_\mu G_\nu - \partial_\nu G_\mu + g_s[G_\mu, G_\nu] . \quad \left( G_\mu = -\frac{i}{2}\lambda^a G_\mu^a \right) \quad (29.14)$$

Note that the color transformations are diagonal in the flavor space (color and flavor transformations do not mix) and the coupling constant  $g_s$  is flavor-independent.

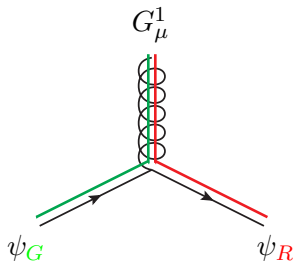
According to the rules of QFT and the associated Feynman diagrams, the  $F^2$  term in the QCD Lagrangian (29.7) gives rise to three basic interactions: a quark may emit (or absorb) a gluon, a gluon may emit (or absorb) a gluon, and two gluons may directly interact, shown schematically in the figure below.





This contrasts with QED, in which only the first kind of interaction occurs, since photons have no charge. Diagrams involving Faddeev-Popov ghosts must be considered too (except in the unitarity gauge).

An example of the color flow for a quark-gluon interaction in color space is given in the figure below, illustrating a quark-quark-gluon vertex before summing/averaging over colors: a gluon in a state represented by  $\lambda^1$  interacts with quarks in the red and green states,  $\psi_R$  and  $\psi_G$ .



$$\begin{aligned}
 &\propto -\frac{i}{2}g_s \quad \bar{\psi}_R \quad \lambda^1 \quad \psi_G = \\
 &= -\frac{i}{2}g_s \quad (1 \ 0 \ 0) \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.
 \end{aligned}$$

Normally, of course, we sum over all the color indices, so this example merely gives a pictorial representation of what one particular (non-zero) term in the color sum looks like.

QCD is a gauge theory and to carry out the quantization gauge fixing is necessary. It means that the QCD Lagrangian given in (29.7) has to be supplemented by a gauge-fixing term and, eventually, by a ghost Lagrangian (in the non-ghost-free gauges, e.g. in the covariant gauge)

$$\mathcal{L}_{\text{QCD}} \rightarrow \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{g.fix.}} + \mathcal{L}_{\text{ghost}} . \tag{29.15}$$

We shall not address this (standard) issue explicitly, since it is not important for the discussion which will be carried out below.

Finally, note that the Lagrangian of QCD contains in addition the so-called  $\theta$ -term

$$\mathcal{L}_\theta = -\frac{\theta}{32\pi^2} \epsilon^{\mu\nu\alpha\beta} F_{\mu\nu}^a F_{\alpha\beta}^a , \tag{29.16}$$

which is related to the axial anomaly and leads to  $CP$  violation in QCD. The physical value of the  $\theta$ -parameter is, however, very small. The necessity to explain why it should be naturally small, constitutes the essence of the so-called *strong CP problem* (see the next lecture).

### 29.3 Color Factors

In QCD calculations we often encounter the some color  $SU(3)$  theoretical factors. Typically, we do not measure color in the final state – instead we average over all possible incoming colors and sum over all possible outgoing ones, wherefore QCD scattering amplitudes (squared) in practice always contain sums over quark fields contracted with Gell-Mann matrices. These contractions in turn produce traces which yield the *color factors* that are associated to each QCD process, and which basically count the number of *paths through color space*,  $N_c$ , that the process at hand can take. Let us consider some processes with different  $N_c$ , taking account of which is important in QCD calculations.

A very simple example of a color factor is given by the decay process

$$Z \rightarrow q\bar{q} . \quad (29.17)$$

This vertex contains a simple  $\delta_{ij}$  in color space; the outgoing quark and anti-quark must have identical (anti-)colors. Squaring the corresponding matrix element and summing over final-state colors yields,

$$e^+e^- \rightarrow Z \rightarrow q\bar{q} \Rightarrow \sum_{\text{colors}} |M|^2 \propto \delta_{ij}\delta_{ji} = \text{Tr}\{\delta\} = N_c = 3 , \quad (29.18)$$

since  $i$  and  $j$  are quark (i.e. 3-dimensional fundamental) indices. This factor corresponds directly to the 3 different paths through color space that the process at hand can take; the produced quarks can be red, green, or blue.

A next-to-simplest example is given by

$$q\bar{q} \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^- , \quad (29.19)$$

which is just a crossing of the previous one. By crossing symmetry, the squared matrix element, including the color factor, is exactly the same as before, but since the quarks are here incoming, we must average rather than sum over their colors, leading to

$$q\bar{q} \rightarrow Z \rightarrow e^+e^- \Rightarrow \frac{1}{9} \sum_{\text{colors}} |M|^2 \propto \frac{1}{9} \delta_{ij}\delta_{ji} = \frac{1}{9} \text{Tr}\{\delta\} = \frac{1}{3} , \quad (29.20)$$

where the color factor now expresses a suppression which can be interpreted as due to the fact that only quarks of matching colors are able to collide and

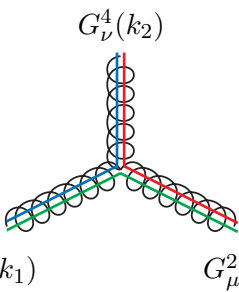
produce a  $Z$  boson. The chance that a quark and an antiquark picked at random from the colliding hadrons have matching colors is  $1/N_c$ .

Similarly,

$$\ell q \rightarrow \ell q \quad (29.21)$$

via  $t$ -channel photon exchange (usually called deep inelastic scattering with 'deep' referring to a large virtuality of the exchanged photon), constitutes yet another crossing of the same basic process. The color factor comes out as unity, i.e. for this case  $N_c = 1$ .

Another example is a gluon self-interaction vertex. Below it is illustrated the interaction between gluons in the states  $\lambda^2$ ,  $\lambda^4$ , and  $\lambda^6$  represented by the structure constant  $f^{246}$ , before summing/averaging over colors.



$$\propto -g_s f^{246} \left[ (k_3 - k_2)^\rho g^{\mu\nu} + (k_2 - k_1)^\mu g^{\nu\rho} + (k_1 - k_3)^\nu g^{\rho\mu} \right]$$

This vertex to be compared with the quark-gluon one given in the previous section. We remind the reader that gauge boson self-interactions are a hallmark of non-Abelian theories and that their presence leads to some of the main differences between QED and QCD.

## 29.4 Renormalization of QCD

The fundamental parameters of QCD are the coupling constant  $g_s$  and the quark masses. All physical observables should be expressed in terms of these parameters. In order to make these observables finite, the bare parameters that enter the Lagrangian should be renormalized. QCD is a renormalizable theory. This means that it suffices to renormalize a finite number of parameters in the Lagrangian, as well as the fields, in order to make all physical observables finite.

### 29.4.1 The Example from QED

As a simple example, we first consider charge renormalization in QED, which is an Abelian counterpart of QCD. The renormalized coupling in QED,  $e_r(\mu)$  (with the scale  $\mu$ ), obeys the Renormalization Group (RG) equation

$$\begin{aligned}\mu \frac{de_r}{d\mu} &= \beta(e_r), \\ \beta(e) &= \frac{e^3}{12\pi^2} + O(e^5).\end{aligned}\tag{29.22}$$

As we see, the coefficient at lowest-order is positive, so the charge grows with the increase of  $\mu$  (at least, for small values of charge, where lowest-order perturbation theory is applicable). The solution of the lowest-order RG equation is given by

$$\frac{1}{e_r^2(\mu)} - \frac{1}{e_r^2(\mu_0)} = -\frac{1}{6\pi^2} \ln \frac{\mu}{\mu_0}.\tag{29.23}$$

Here,  $\mu_0$  is some scale, where the initial condition for the RG equation is set.

The solution of the differential equation can be written in a form which makes the presence of a (perturbative) Landau pole explicitly,

$$e_r^2(\mu) = \frac{e_r^2(\mu_0)}{1 - \frac{e_r^2(\mu_0)}{6\pi^2} \ln \frac{\mu}{\mu_0}} = \frac{6\pi^2}{\ln \frac{\Lambda_{\text{QED}}}{\mu}}.\tag{29.24}$$

Here we have defined the dimensionful quantity  $\Lambda_{\text{QED}}$  according to

$$\Lambda_{\text{QED}} = \mu_0 e^{6\pi^2/e_r^2(\mu_0)}.\tag{29.25}$$

It can be straightforwardly checked that  $\Lambda_{\text{QED}}$  is RG-invariant

$$\mu_0 \frac{d\Lambda_{\text{QED}}}{d\mu_0} = \Lambda_{\text{QED}} - \mu_0^2 \frac{12\pi^2}{e_r^3(\mu_0)} \frac{de_r(\mu_0)}{d\mu_0} e^{6\pi^2/e_r^2(\mu_0)} = 0.\tag{29.26}$$

This result is very remarkable. In fact, it shows that a scale-dependent dimensionless coupling constant  $e_r(\mu_0)$  can be traded versus the RG-invariant dimensionful coupling constant  $\Lambda_{\text{QED}}$ . This becomes clear from the last equation in (29.24), which does not contain the coupling constant at all. Putting it differently,  $\Lambda_{\text{QED}}$  plays the role of a coupling constant. This phenomenon goes under the name of *dimensional transmutation*.

QED is not a confining theory – electrons, positrons and photons exist in the asymptotic state. The only relevant scale in QED is the electron mass. Setting

$$\mu_0 = m_e \simeq 0.5 \text{ MeV} , \quad \frac{e_r^2(\mu_0)}{4\pi} \simeq \frac{1}{137} , \quad (29.27)$$

one may estimate the size of  $\Lambda_{\text{QED}}$ . This estimate results in

$$\Lambda_{\text{QED}} \simeq 10^{277} \text{ GeV} , \quad (29.28)$$

which is much larger than the Planck mass,  $M_{\text{Planck}} \simeq 10^{19} \text{ GeV}$ . Hence, to a very good approximation, QED is a local theory. Even if a Landau pole appears at  $\mu = \Lambda_{\text{QED}}$ , this is very far away from any measurable energy.

### 29.4.2 RG in QCD

One can repeat the above arguments in case of QCD. The RG equation for the coupling constant looks as follows

$$\mu \frac{dg_r(\mu)}{d\mu} = \beta(g_r) , \quad (29.29)$$

where the perturbative expansion of the  $\beta$ -function now starts with the term with a *negative* coefficient

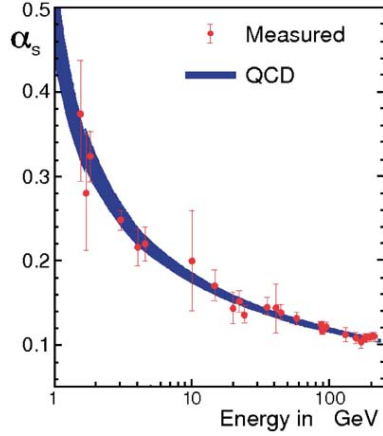
$$\begin{aligned} \beta(g) &= -\beta_0 \frac{g^3}{16\pi^2} - \beta_1 \frac{g^5}{(16\pi^2)^2} + O(g^7) , \\ \beta_0 &= \frac{11}{3} N_c - \frac{2}{3} N_f , \\ \beta_1 &= \frac{34}{3} N_c^2 - \frac{10}{3} N_c N_f - \frac{(N_c^2 - 1)N_f}{N_c} . \end{aligned} \quad (29.30)$$

Here,  $N_c = 3$  and  $N_f$  stand for the number of colors and number of flavors, respectively. One sees that, if  $N_f$  is sufficiently small, then  $\beta_0 > 0$  and the renormalized coupling constant  $g_r(\mu)$  *decreases* as  $\mu \rightarrow \infty$ . This behavior amounts to the *asymptotic freedom* of the theory and is inherent to *non-Abelian* theories. As we have seen, QED, which is an Abelian theory, does not possess the property of asymptotic freedom.

Below you can see dependence of the strong coupling constant

$$\alpha_s(E) = \frac{g_r^2(E)}{4\pi} . \quad (29.31)$$

on the energy.



At small energies, the coupling constant of QCD grows and the theory becomes non-perturbative. Moreover, QCD is a confining theory: instead of quarks and gluons, only colorless bound states thereof are observed in experiment. A typical mass scale of such states is of order of the nucleon mass, i.e. around 1 GeV.

The analog of the  $\Lambda$ -parameter in QCD *to all orders in the strong coupling constant* is defined by

$$\Lambda_{\text{QCD}} = \mu_0 \left( \frac{\beta_0 g_r^2}{16\pi^2} \right)^{-\beta_1/2\beta_0^2} e^{-\frac{16\pi^2}{2\beta_0 g_r^2} - \int dg' \left( \frac{1}{\beta(g')} + \frac{16\pi^2}{\beta_0 g'^3} - \frac{\beta_1}{\beta_0^2 g'} \right)}. \quad (29.32)$$

Note that (29.26) is obtained from (29.32) in the lowest order in perturbation theory, if the pertinent value for  $\beta_0$  in QED is used. Note also that the subtraction in (29.32) in the integral regularizes the integrand at  $g' = 0$ , where  $\beta(g')$  has a cubic singularity. The other factors, which enter the expression in (29.32), are chosen so that the quantity  $\Lambda_{\text{QCD}}$  is RG-invariant to all orders in perturbation theory

$$\mu_0 \frac{d\Lambda_{\text{QCD}}}{d\mu_0} = 0. \quad (29.33)$$

Using now the lowest-order result, assuming  $\mu_0 \simeq 1$  GeV (the typical mass of the QCD bound states), setting  $N_f = 3$  according to the number of quark flavors lighter than 1 GeV, and taking  $\alpha_s(E = 1 \text{ GeV}) \simeq 0.4$ , we get a rough estimate

$$\Lambda_{\text{QCD}} \simeq 200 \text{ MeV}. \quad (29.34)$$

Hence, in difference to QED, QCD is essentially non-local at low energies.

One may wonder, why, in analogy to QED, one did not use quark masses (the only parameters with the dimension of mass in the Lagrangian), in order to fix the QCD scale. In QCD, the renormalized masses obey RG equations

$$\mu \frac{dm_r^f(\mu)}{d\mu} = -\gamma(g_r)m_r^f(\mu), \quad (f = u, d, s, c, b, t) \quad (29.35)$$

where the  $\gamma$ -function has the following perturbative expansion

$$\begin{aligned} \gamma(g) &= \gamma_0 \frac{g^2}{16\pi^2} + \gamma_1 \frac{g^4}{(16\pi^2)^2} + O(g^6), \\ \gamma_0 &= \frac{3(N_c^2 - 1)}{N_c}, \\ \gamma_1 &= 3 \left( \frac{N_c^2 - 1}{2N_c} \right)^2 + \frac{97}{6} (N_c^2 - 1) - \frac{5N_c(N_c^2 - 1)}{3N_c}. \end{aligned} \quad (29.36)$$

Note that the  $\gamma$ -function does not depend on the quark flavor that reflects the fact that strong interactions are flavor-blind. Consequently, the renormalized quark masses run with the scale  $\mu$ , but the ratios  $m_r^f(\mu)/m_r^{f'}(\mu)$  are scale-independent in QCD.

The range of the quark mass values varies from a few MeV ( $u$  and  $d$  quarks) to  $\simeq 100$  MeV ( $s$ -quark) and higher (according to the commonly used convention, we fix the scale  $\mu = 2$  GeV). Since the quarks of higher masses can not play any role in low-energy physics we are primarily interested in, below we shall concentrate on three light flavors only.

The answer to the question that was posed above lies in a peculiar character of the QCD spectrum. Namely, the light quark masses at the scale of order of few GeV are much smaller than the typical hadron masses (say, the nucleon mass). So, in the first approximation, one may put all quark masses to zero. The low-energy spectrum does not change dramatically – say, the nucleon mass shifts downwards by approximately 100 MeV. Consequently, the low-energy hadron spectrum is mainly determined through the interaction energy of the gluon field inside the hadrons and not the quark masses. Reversing the argument, it is natural to use the low-lying hadron masses to fix the strength of gluonic interactions, which is encoded in the quantity  $\Lambda_{\text{QCD}}$ .

There is however one important exception from this picture, related to the phenomenon of the *chiral SSB in QCD*, which manifests itself in the emergence of an octet of pseudoscalar Goldstone bosons: pions, kaons and eta.

The masses of these particles are proportional to the quark masses and vanish in the chiral limit,  $m_r^f \rightarrow 0$ .

To summarize, dimensional transmutation in QCD introduces the RG invariant scale  $\Lambda_{\text{QCD}}$  in QCD which is a substitute of the scale-dependent, dimensionful coupling constant  $g_r(\mu)$  and is determined by the low-lying hadronic spectrum in QCD. A typical size of the hadronic scale is given by the nucleon mass (around 1 GeV) that leads to  $\Lambda_{\text{QCD}} \simeq 200$  MeV. Only the octet of Goldstone bosons is much lighter, owing to the (approximate) chiral symmetry of QCD.

QCD as a field theory becomes non-local below the hadronic scale  $\sim 1$  GeV. Quarks and gluons are not the appropriate degrees of freedom any more, paving the way for the description in terms of the hadronic fields. Since the Goldstone bosons are the only hadrons whose masses are far below the hadronic scale, the effective theory of QCD at low energies is primarily the theory of Goldstone bosons, and incorporates the restrictions, stemming from chiral symmetry, in an essential manner.

**Exercise 29.1:** Explain why each of the following particles cannot exist according to the quark model.

- (a) A baryon of spin 1;
- (b) An antibaryon of electric charge +2;
- (c) A meson with charge +1 and strangeness  $-1$ ;
- (d) A meson with opposite signs of charm and strangeness.

**Exercise 29.2:** Let  $\psi^T = (u, d, s)$  be an  $SU(3)$  tripler. Decompose the product  $\psi_i^* \psi_j$  (where  $i, j = 1, 2, 3$ ) into irreducible representations of  $SU(3)$ .

**Exercise 29.3:** Find the overall "color factor" for  $qq \rightarrow qq$  if QCD corresponded to a  $SU(2)$  color symmetry.

**Exercise 29.4:** Calculate the octet  $q\bar{q}$  color factor using the states: (a)  $B\bar{G}$ ; (b)  $(R\bar{R} - B\bar{B})/\sqrt{2}$ ; (c)  $(R\bar{R} + B\bar{B} - 2G\bar{G})/\sqrt{6}$ .

**Exercise 29.5:** Calculate  $\alpha_s$  at 10 and 100 GeV.

**Exercise 29.6:** From the expression for the running of  $\alpha_s$  with  $N_f = 3$ , determine the value of energy at which  $\alpha_s$  appears to become infinite.



## Chapter 30

# Symmetry Breaking Patterns in QCD

### 30.1 Chiral Symmetry in QCD

In this section, we consider the implications of chiral symmetry in QCD. In the following, the gluonic part of the QCD Lagrangian will not play a role and we shall concentrate on the fermionic part which is given by

$$\mathcal{L}_F = \bar{\psi}(i\gamma^\mu D_\mu - \mathcal{M})\psi . \quad (30.1)$$

In addition, we shall restrict ourselves to the light flavors  $N_f = 3$

$$\mathcal{M} = \text{diag}(m_u, m_d, m_s) , \quad \psi(x) = \begin{pmatrix} u(x) \\ d(x) \\ s(x) \end{pmatrix} , \quad (30.2)$$

since, at low energies, the heavy quark degrees of freedom decouple from the theory.

At the next step, let us decompose the fermion field  $\psi(x)$  into the left- and right-hand components, according to

$$\begin{aligned} \psi_L(x) &= \frac{1}{2}(1 - \gamma_5)\psi(x) , & \psi_R(x) &= \frac{1}{2}(1 + \gamma_5)\psi(x) , \\ \psi(x) &= \psi_L(x) + \psi_R(x) . \end{aligned} \quad (30.3)$$

The fermionic part of the Lagrangian in terms of the left- and right-hand fields has the form:

$$\mathcal{L}_F = \bar{\psi}_L(i\gamma^\mu D_\mu)\psi_L + \bar{\psi}_R(i\gamma^\mu D_\mu)\psi_R - \bar{\psi}_L\mathcal{M}\psi_R - \bar{\psi}_R\mathcal{M}\psi_L . \quad (30.4)$$

From this expression it is immediately clear that, if the quarks are massless,  $\mathcal{M} = 0$ , then  $\mathcal{L}_F$  is invariant under the *global* group  $U_L(N_f) \times U_R(N_f)$

$$\begin{aligned} \psi_L(x) &\mapsto g_L\psi_L(x) , & g_L &\in U_L(N_f) , \\ \psi_R(x) &\mapsto g_R\psi_R(x) , & g_R &\in U_R(N_f) . \end{aligned} \quad (30.5)$$

Note that the transformations from the group  $U_L(N_f) \times U_R(N_f)$  do not touch the gluon field, which is flavor blind. Thus, the gluonic (and ghost) part of the Lagrangian is trivially invariant under the above transformations.

The symmetry of the Lagrangian results in the conservation of the left- and right-hand currents. Non-singlet currents are defined as

$$\begin{aligned} L_\mu^i &= \bar{\psi}_L\gamma_\mu T^i\psi_L = \bar{\psi}\gamma_\mu\frac{1}{2}(1 - \gamma_5)T^i\psi , \\ R_\mu^i &= \bar{\psi}_R\gamma_\mu T^i\psi_R = \bar{\psi}\gamma_\mu\frac{1}{2}(1 + \gamma_5)T^i\psi , \end{aligned} \quad (30.6)$$

where  $T^i$  ( $i = 1, \dots, N_f^2 - 1$ ) are the generators of the  $SU(N_f)$  group. The singlet currents take the form

$$\begin{aligned} L_\mu^0 &= \bar{\psi}\gamma_\mu(1 - \gamma_5)\psi , \\ R_\mu^0 &= \bar{\psi}\gamma_\mu(1 + \gamma_5)\psi , \end{aligned} \quad (30.7)$$

If  $\mathcal{M} = 0$ , then

$$\partial^\mu L_\mu^i = \partial^\mu R_\mu^i = \partial^\mu L_\mu^0 = \partial^\mu R_\mu^0 = 0 , \quad (30.8)$$

for a moment, we neglect the anomalies.

The vector- and axial-vector currents are defined by

$$\begin{aligned} V_\mu^i &= \frac{1}{2}(R_\mu^i + L_\mu^i) , & A_\mu^i &= \frac{1}{2}(R_\mu^i - L_\mu^i) , \\ V_\mu^0 &= \frac{1}{2}(R_\mu^0 + L_\mu^0) , & A_\mu^0 &= \frac{1}{2}(R_\mu^0 - L_\mu^0) . \end{aligned} \quad (30.9)$$

If  $\mathcal{M} = 0$ , these currents are conserved along with the left- and right-hand currents.

Assume now that the quark masses are nonzero. Calculating the derivative

$$\partial^\mu V_\mu^i = \bar{\psi} \frac{1}{2} T^i \gamma_\mu \partial^\mu \psi + (\partial^\mu \bar{\psi}) \gamma^\mu \frac{1}{2} T^i \psi , \quad (30.10)$$

and using equation of motion

$$i\gamma^\mu \left( \partial_\mu - \frac{i}{2} \lambda^a D_\mu^a \right) \psi - \mathcal{M}\psi = 0 , \quad (30.11)$$

we obtain

$$\partial^\mu V_\mu^i = -i\bar{\psi} \frac{1}{2} [T^a, \mathcal{M}] \psi . \quad (30.12)$$

Analogously,

$$\partial^\mu A_\mu^i = -i\bar{\psi} \frac{1}{2} \{T^a, \mathcal{M}\} \gamma_5 \psi . \quad (30.13)$$

This means that the non-singlet vector current is broken by the *differences in the quark masses* ( $m_u = m_d = m_s \neq 0$ ) – it stays conserved. On the contrary, the axial-vector current is violated by the *quark masses itself* – if the quark masses are nonzero, then it is not conserved.

An important remark is in order. To calculate the divergences of the currents, one has used the equations of motion. However, in field theory, the products of different operators that enter equations of motion, are singular, and such calculations, in general, cannot be done without a specific care. Carrying out calculations with more mathematical rigor (see below), it turns out that the divergence of the singlet axial-vector current contains a term that does not vanish even in the limit  $\mathcal{M} \rightarrow 0$

$$\partial^\mu A_\mu^0 = 2i\bar{\psi} \mathcal{M} \gamma_5 \psi - \frac{N_f}{(4\pi)^2} \epsilon_{\mu\nu\alpha\beta} \text{tr}_c \left( F^{\mu\nu} F^{\alpha\beta} \right) , \quad (30.14)$$

where  $\text{tr}_c$  denotes the trace over color indices. The last term in the above expression goes under the name of *axial anomaly*.

Finally, note that the singlet vector current  $V_\mu^0$  is always conserved, that corresponds to the conservation of the number of quarks minus the number of the antiquarks. The zeroth component of this current is proportional to the baryon number density.

### 30.1.1 Chiral SSB

As discussed in the previous section, owing to the anomaly, the symmetry of the Lagrangian (massless case) is broken down to

$$U_L(N_f) \times U_R(N_f) \rightarrow SU_L(N_f) \times SU_R(N_f) \times U_V(1) \quad (30.15)$$

(Another diagonal subgroup,  $U_A(1)$ , is violated by the anomaly). Next, we wish to answer the question: what is the symmetry of Nature, which is described by this Lagrangian?

In the beginning, let us restrict ourselves first to the subgroup  $SU_L(N_f) \times SU_R(N_f)$  and consider the conserved charges

$$\begin{aligned} Q_V^i(t) &= \int d^3\mathbf{x} V_0^i(\mathbf{x}, t), \quad (i = 1, \dots, N_f^2 - 1) \\ Q_A^i(t) &= \int d^3\mathbf{x} A_0^i(\mathbf{x}, t). \end{aligned} \quad (30.16)$$

In the massless case, we have

$$\dot{Q}_{V,A}^i(t) = i[H, Q_{V,A}^i(t)] = 0. \quad (30.17)$$

Here,  $H$  denotes the Hamiltonian of the system. Since the  $Q_{V,A}^i(t)$  do not depend on the time, we further drop the argument  $t$  in this expression.

Consider now a generic conserved charge  $Q^i(t)$  (denotes either  $Q_V^i$  or  $Q_A^i$ ), which commutes with the Hamiltonian. It is possible to have two different *modes of realization* of the above symmetry, depending on whether the corresponding charges annihilate the vacuum:

- Wigner-Weyl mode:  $Q^i(t)|0\rangle = 0$ ;
- Nambu-Goldstone mode:  $Q^i(t)|0\rangle \neq 0$ .

Consider first the symmetry realization in the Wigner-Weyl mode. Assume that there is a mass gap in the theory, and consider the one-particle state with a mass  $m$  in the rest-frame. Since  $Q^i$  commutes with the Hamiltonian,

$$HQ^i|\mathbf{p} = \mathbf{0}\rangle = Q^iH|\mathbf{p} = \mathbf{0}\rangle = mQ^i|\mathbf{p} = \mathbf{0}\rangle. \quad (30.18)$$

This means that  $Q^i|\mathbf{p} = \mathbf{0}\rangle$  is also a one-particle state with the same mass – in other words, the massive particles form the *symmetry multiplets*, where

the individual states are labeled by some index  $\alpha$ . Introducing the creation operator, one may write

$$\begin{aligned} Q^i a_\alpha^\dagger(\mathbf{p} = \mathbf{0})|0\rangle &= [Q^i, a_\alpha^\dagger(\mathbf{p} = \mathbf{0})]|0\rangle + a_\alpha^\dagger(\mathbf{p} = \mathbf{0})Q^i|0\rangle = \\ &= C_{\alpha\beta}^i a_\alpha^\dagger(\mathbf{p} = \mathbf{0})|0\rangle . \end{aligned} \quad (30.19)$$

Here, we have used the condition

$$Q^i|0\rangle = 0 \quad (30.20)$$

and the fact that, acting with an operator  $Q^i$  on any member of a multiplet, we can only get a linear combination (with some coefficients  $C_{\alpha\beta}^i$ ) of all states in this multiplet but nothing else. From this we finally conclude that in the Wigner-Weyl mode the symmetry is realized linearly by an operator

$$U(\omega) = e^{i\omega\mathbf{Q}} , \quad (30.21)$$

where the components of the vector  $\omega$  are the parameters of a symmetry transformation. In a given multiplet, where the individual states form a basis of an (irreducible) representation, the matrices of this irreducible representation are given by

$$T(\omega)_{\alpha\beta} = e^{i(\omega^i C^i)_{\alpha\beta}} . \quad (30.22)$$

In case of the chiral symmetry, the implications are far-reaching. Assume that chiral symmetry in QCD is realized in the Wigner-Weyl mode. Since under parity  $Q_A^i \mapsto -Q_A^i$ , one should have *the multiplets with opposite parity and with the same mass* in QCD. If the quark masses are turned on, the masses in these multiplets would become slightly different but, in case of small quark masses, the pattern should be easily recognizable. This, however, is not what one observes in experiment. For example, the lightest pseudoscalar bosons (pions) have the mass around 140 MeV, whereas the lightest scalar boson has the mass around 450 MeV and the width above 500 MeV. The similar picture emerges for other multiplets. For example, the lowest excited state of the nucleon with negative parity, the  $S_{11}(1535)$  with the quantum numbers  $1/2^-$ , lies approximately 600 MeV above the nucleon ground state. It is evident that the symmetry pattern of the QCD spectrum at low energies cannot be consistently described within this picture.

Let us now consider the alternative picture, where the charges  $Q^i$  do not annihilate the vacuum. This is equivalent to the statement that there exists

some operator  $B$ , for which

$$\begin{aligned}\langle 0|BQ^i|0\rangle &= \sum_n \int d\mathbf{x} \langle 0|B|n\rangle \langle n|j_0^i(\mathbf{x}, t)|0\rangle = \\ &= \sum_n (2\pi)^3 \delta^3(\mathbf{P}_n) e^{iE_n t} \langle 0|B|n\rangle \langle n|j_0^i(0)|0\rangle \neq 0, \quad (30.23)\end{aligned}$$

where  $j_\mu^i(x)$  denotes the current associated with the conserved charge  $Q^i$ , and  $P_{n\mu} = (E_n, \mathbf{P}_n)$  is the four-momentum of the state  $|n\rangle$ .

On the other hand, since  $Q^i$  is conserved, one may write

$$\frac{d}{dt} \langle 0|BQ^i|0\rangle = \sum_n (2\pi)^3 \delta^3(\mathbf{P}_n) iE_n e^{iE_n t} \langle 0|B|n\rangle \langle n|j_0^i(0)|0\rangle = 0. \quad (30.24)$$

This is possible, only if the states  $|n\rangle$ , which are produced by acting of  $Q^i$  on the vacuum state, are massless:

$$E_n(\mathbf{P}_n = \mathbf{0}) = 0. \quad (30.25)$$

Consequently, in this case the massless *Nambu-Goldstone bosons* emerge in the spectrum. Each symmetry generator  $Q^i$ , which does not annihilate the vacuum, gives rise to one massless boson. One speaks about the *chiral SSB*.

Note that, if the symmetry is spontaneously broken, a special care should be taken, working out the infinite-volume limit in the theory, which is defined in a finite volume. For example, since the massless Goldstone bosons have infinite correlation length, they cannot emerge in a finite volume. Consequently, a correct treatment of the SSB implies considering *first* the infinite-volume limit and *then* the limit where the explicit symmetry breaking parameters (e.g., the quark masses) tend to zero. These limits cannot be interchanged.

In QCD, the symmetry is spontaneously broken for the axial charges

$$Q_A^i|0\rangle \neq 0. \quad (30.26)$$

Instead of parity doublets, eight Goldstone bosons (pions, kaons and eta) arise in result of spontaneous symmetry breaking. These (pseudoscalar) bosons would have exactly zero mass in the chiral limit

$$m_u, m_d, m_s \rightarrow 0. \quad (30.27)$$

The measured masses emerge in result of the explicit symmetry breaking by quark masses.

The SSB is signaled, for example, by the fact that

$$\langle 0 | \bar{\psi}(0) \psi(0) | 0 \rangle = \langle 0 | \bar{\psi}_L(0) \psi_R(0) + \bar{\psi}_R(0) \psi_L(0) | 0 \rangle \neq 0 . \quad (30.28)$$

This quantity is invariant under the vector sub-group  $g_L = g_R$ , but not under general chiral transformations. Consequently, having this quantity non-vanishing is a sufficient but not a necessary condition of the spontaneous chiral symmetry breaking.

On the other hand, if the axial-vector symmetry is spontaneously broken

$$Q_A^i | 0 \rangle \neq 0 , \quad (30.29)$$

then, as seen from (30.23), the matrix elements of the axial-vector current between the vacuum and the one-Goldstone-boson state does not vanish. Namely, in QCD, a necessary and sufficient condition for the SSB is the non-vanishing value of the pion decay constant in the chiral limit. This quantity is defined through the matrix element of the axial-vector current between the vacuum and the one-pion state

$$\langle 0 | A_\mu^k(x) | \pi^i(q) \rangle = i q_\mu \delta_{ik} F_\pi e^{-iqx} , \quad F_0 = \lim_{\mathcal{M} \rightarrow 0} F_\pi . \quad (30.30)$$

The quantity  $F_0$  represents the true *order parameter* in QCD, with  $F_0 \neq 0$  corresponding to the spontaneously broken phase. The *quark condensate* defined by (30.28), if non-vanishing, also serves as an order parameter. However, vanishing of the quark condensate does not necessarily mean the absence of the chiral SSB.

If the singlet axial-vector current were also conserved in the chiral limit, there would be an associated ninth Goldstone boson in the spectrum. However, the conservation of the singlet current is broken by anomaly, and the would-be Goldstone boson  $\eta'$  has a mass around 1 GeV.

Finally, note that the vector symmetry in QCD, unlike the axial-vector symmetry, is not spontaneously broken

$$Q_V^i | 0 \rangle = 0 . \quad (30.31)$$

This follows from the Vafa-Witten theorem.

## 30.2 Hadron Masses

We assume that the quark masses are small as compared to the QCD scale and calculate the first-order perturbation to the hadron masses in the chiral limit. The QCD Hamiltonian can be written in the following form

$$\begin{aligned} H_{\text{QCD}} &= H_0 + H_I, \\ H_I &= \int d^3\mathbf{x} [m_u \bar{u}(x)u(x) + m_d \bar{d}(x)d(x) + m_s \bar{s}(x)s(x)] \end{aligned} \quad (30.32)$$

where  $H_0$  denotes the Hamiltonian in the chiral limit. Let now  $|\mathbf{p}, n\rangle$  be the covariantly normalized 1-particle eigenstates states of  $H_0$  with a mass  $M_n$

$$H_0|\mathbf{p}, n\rangle = E_n(\mathbf{p})|\mathbf{p}, n\rangle, \quad (\langle \mathbf{p}', n|\mathbf{p}, n\rangle = (2\pi)^3 \delta^3(\mathbf{p}' - \mathbf{p}) 2E_n(\mathbf{p})) \quad (30.33)$$

where  $E_n(\mathbf{p}) = \sqrt{M_n^2 + \mathbf{p}^2}$ .

Calculating the energy shift at the first order, we have

$$\begin{aligned} \langle \mathbf{p}', n|H_I|\mathbf{p}, n\rangle &= (2\pi)^3 \delta^3(\mathbf{p}' - \mathbf{p}) \langle \mathbf{p}', n|m_u \bar{u}u + m_d \bar{d}d + m_s \bar{s}s|\mathbf{p}, n\rangle = \\ &= \delta E_n(\mathbf{p}) (2\pi)^3 \delta^3(\mathbf{p}' - \mathbf{p}) 2E_n(\mathbf{p}). \end{aligned} \quad (30.34)$$

From the above relation, one immediately obtains

$$2E_n(\mathbf{p})\delta E_n(\mathbf{p}) = \delta M_n^2 = \langle \mathbf{p}, n|m_u \bar{u}u + m_d \bar{d}d + m_s \bar{s}s|\mathbf{p}, n\rangle. \quad (30.35)$$

If the mass  $M_n$  does not vanish in the chiral limit, one may write

$$\begin{aligned} M_n^2 &= M_n^0 + m_u B_n^u + m_d B_n^d + m_s B_n^s + \dots, \\ B_n^f &= \lim_{\mathcal{M} \rightarrow 0} \langle \mathbf{p}, n|\bar{\psi}_f \psi_f|\mathbf{p}, n\rangle. \quad (f = u, d, s) \end{aligned} \quad (30.36)$$

Note that the quantities  $M_n^0$  and  $B_n^f$  depend only on a single parameter  $\Lambda_{\text{QCD}}$ .

The equation (30.36) can be presented in a form linear in hadron masses

$$M_n = M_n^0 + m_u \tilde{B}_n^u + m_d \tilde{B}_n^d + m_s \tilde{B}_n^s + \dots. \quad (30.37)$$

The relation between the parameters  $B_n^f$  and  $\tilde{B}_n^f$  can be easily established. This procedure, however, fails, if we are dealing with the Goldstone bosons,



for which the quantity  $M_n^0$  vanishes. In this case, one has to use the quadratic form given in (30.36).

Next, one may use the fact that in the chiral limit QCD has exact  $SU_V(3)$  symmetry. Writing

$$\begin{aligned}\bar{u}u &= u^0 + \frac{1}{2}u^3 + \frac{1}{2\sqrt{3}}u^8, \\ \bar{d}d &= u^0 - \frac{1}{2}u^3 + \frac{1}{2\sqrt{3}}u^8, \\ \bar{s}s &= u^0 - \frac{1}{\sqrt{3}}u^8,\end{aligned}\tag{30.38}$$

where  $u^i = \bar{\psi}\lambda^i\psi$  ( $i = 1, \dots, 8$ ) and  $u^0 = (\bar{u}u + \bar{d}d + \bar{s}s)/3$ , it is straightforward to check that  $u^i$  transforms like a  $SU_V(3)$ -octet, whereas  $u^0$  is a scalar. Since the states  $|\mathbf{p}, n\rangle$ , which are the eigenstates of the QCD Hamiltonian *in the chiral limit*, also form the  $SU_V(3)$ -multiplets, one may use the Wigner-Eckart theorem, to relate the matrix elements  $B_n^f$  with a different  $n$  and  $f$ . For example, in case of the Goldstone bosons the result is particularly simple

$$\begin{aligned}M_\pi^2 &= 2\hat{m}B_0 + \dots, \\ M_K^2 &= (\hat{m} + m_s)B_0 + \dots, \\ M_\eta^2 &= \frac{2}{3}(\hat{m} + 2m_s)B_0 + \dots,\end{aligned}\tag{30.39}$$

where we additionally assumed isospin symmetry to be exact  $m_u = m_d = \hat{m}$ . The ellipses stand for the higher-order terms in the quark masses. As seen, all masses in the multiplet are expressed in terms of a single matrix element

$$B_0 = \langle \pi^0 | \bar{u}u | \pi^0 \rangle.\tag{30.40}$$

This fact is a consequence of an exact  $SU_V(3)$  symmetry of the theory in the chiral limit.

Finally, it is immediately seen that the famous Gell-Mann-Okubo formula is readily reproduced at the lowest order in the quark mass expansion

$$4M_K^4 - M_\pi^2 - 3M_\eta^2 = 0.\tag{30.41}$$

This relation is fulfilled in Nature within a few percent accuracy.

### 30.3 The Example of $\sigma$ -model

The non-linear  $\sigma$ -model served as the dominant prototype of SSB of  $O(4)$  down to  $O(3)$ : the three axial generators broken are the simplest manifestation of chiral symmetry breaking, the surviving unbroken  $O(3)$  representing isospin. The model was introduced by Gell-Mann and Lévy (1960), who named it after a field corresponding to a spinless meson called  $\sigma$  in their model, a scalar meson introduced earlier by Schwinger.

#### 30.3.1 Lagrangian of the $\sigma$ -model

In this section, we construct a simple renormalizable model, which follows the same symmetry pattern as QCD at low energies. Although this is not a full-fledged low-energy effective field theory of QCD, since it does not contain *all* operators with the pertinent symmetries, it is a first (very instructive) step towards the construction of such an effective field theory.

In the following, for simplicity, we restrict ourselves to the two lightest flavors  $N_f = 2$ . Consider fields forming a 4-vector  $(\sigma, \boldsymbol{\pi})$ , where  $\sigma$  and  $\boldsymbol{\pi}$  are a scalar field and a triplet of the pseudoscalar fields, transforming as quark bilinears  $\bar{\psi}\psi$  and  $\bar{\psi}i\gamma^5\boldsymbol{\tau}\psi$ , respectively (here,  $\boldsymbol{\tau}$  stands for the Pauli isospin matrices).

It is useful to introduce a  $2 \times 2$  matrix

$$\Sigma = \sigma \mathbf{1} + i\boldsymbol{\pi}\boldsymbol{\tau} . \quad (30.42)$$

Under the  $SU_L(2) \times SU_R(2)$  transformations, the matrix  $\Sigma$  transforms as

$$\Sigma \mapsto g_R \Sigma g_L^\dagger . \quad (30.43)$$

Under the infinitesimal transformations

$$\begin{aligned} g_L &= \mathbf{1} + i\boldsymbol{\alpha} - i\boldsymbol{\beta} , & (\boldsymbol{\alpha} = \boldsymbol{\alpha}\boldsymbol{\tau}, \boldsymbol{\beta} = \boldsymbol{\beta}\boldsymbol{\tau}) \\ g_R &= \mathbf{1} + i\boldsymbol{\alpha} + i\boldsymbol{\beta} , \end{aligned} \quad (30.44)$$

the fields transform as

$$\begin{aligned} \sigma &\mapsto \sigma - 2\beta^i \pi^i , \\ \pi^i &\mapsto \pi^i - 2\epsilon^{ijk} \alpha^j \pi^k + 2\beta^i \sigma . \end{aligned} \quad (30.45)$$

The Lagrangian of the linear  $\sigma$ -model is given by

$$\mathcal{L} = \frac{1}{4} \langle \partial^\mu \Sigma \partial_\mu \Sigma^\dagger \rangle - \frac{\mu^2}{4} \langle \Sigma \Sigma^\dagger \rangle - \frac{\lambda}{16} \langle \Sigma \Sigma^\dagger \rangle^2 + \frac{c}{4} \langle \Sigma + \Sigma^\dagger \rangle = \mathcal{L}_0 + c\mathcal{L}_1, \quad (30.46)$$

where  $\langle \dots \rangle$  denotes the trace. It is clear that  $\mathcal{L}_0$  is invariant under the  $SU_L(2) \times SU_R(2)$  group, whereas the term  $c\mathcal{L}_1$  breaks this symmetry down to the diagonal  $SU_V(2)$  subgroup. This term serves as an analog for an explicit chiral symmetry breaking term in QCD, which is proportional to the quark masses.

In terms of the component fields the Lagrangian takes the form

$$\mathcal{L} = \frac{1}{2} (\partial \boldsymbol{\pi})^2 + \frac{1}{2} (\partial \sigma)^2 - \frac{\mu^2}{2} (\boldsymbol{\pi}^2 + \sigma^2) - \frac{\lambda}{4} (\boldsymbol{\pi}^2 + \sigma^2)^2 + c\sigma. \quad (30.47)$$

The vector and axial-vector currents, obtained by using the Noether theorem, are given by

$$\begin{aligned} V_\mu^i &= \epsilon^{ijk} \pi^j \partial_\mu \pi^k, \\ A_\mu^i &= \sigma \partial_\mu \pi^i - \pi^i \partial_\mu \sigma. \end{aligned} \quad (30.48)$$

If  $c \neq 0$ , the axial-vector current is not conserved

$$\partial^\mu A_\mu^i = -c\pi^i. \quad (30.49)$$

As we have learned, in order to establish, how the symmetries are realized in a theory, the study of the symmetries of the Lagrangian is not enough. The symmetry of the ground state in the theory should be studied as well. In the tree approximation, the ground state is obtained by minimizing the classical potential of the fields. Since we do not wish to allow for the spontaneous breaking of parity, one has to assume that the vacuum expectation value of the pseudoscalar field  $\boldsymbol{\pi}$  is zero. Consequently, the potential to be minimized will depend on the field  $\sigma$  alone (for a moment, we consider a case with  $c = 0$ , where symmetry is not *explicitly* broken)

$$V(\sigma) = \frac{\mu^2}{2} \sigma^2 + \frac{\lambda}{4} \sigma^4. \quad (30.50)$$

One has to distinguish between two cases:  $\mu^2 > 0$  and  $\mu^2 < 0$  (we remind the reader that  $\lambda > 0$  always, in order to ensure that the potential is bound from below). In the first case, the minimum of the potential occurs at  $\sigma = 0$ . The

masses of the pion and  $\sigma$  are both equal to  $\mu$ , and the symmetry is realized in the Wigner-Weyl mode. The second case is more interesting. The equation for a local extremum is given by

$$V'(\sigma) = \mu^2\sigma + \lambda\sigma^3 = 0. \quad (30.51)$$

Now the solution  $\sigma = 0$  corresponds to a local maximum of the potential. The minima occur at  $\sigma = \pm v$ , where  $v^2 = -\mu^2/\lambda > 0$ . Introducing the shift  $\sigma \rightarrow \sigma + v$ , so that the new field  $\sigma$  has a vanishing vacuum expectation value  $\langle 0|\sigma|0\rangle = 0$ , and expanding the Lagrangian in the vicinity of a new minimum, one, up to an inessential constant, gets

$$\begin{aligned} \mathcal{L}_0 = & \frac{1}{2}(\partial\boldsymbol{\pi})^2 + \frac{1}{2}(\partial\sigma)^2 - \frac{(\mu^2 + 3\lambda v^2)}{2}\sigma^2 - \frac{(\mu^2 + \lambda v^2)}{2}\boldsymbol{\pi}^2 - \\ & -\lambda v\sigma(\boldsymbol{\pi}^2 + \sigma^2) - \frac{\lambda}{4}(\boldsymbol{\pi}^2 + \sigma^2)^2. \end{aligned} \quad (30.52)$$

From this expression, one can immediately read off the masses

$$\begin{aligned} M_\pi^2 &= \mu^2 + \lambda v^2 = 0, \\ M_\sigma^2 &= \mu^2 + 3\lambda v^2 = -2\mu^2 > 0. \end{aligned} \quad (30.53)$$

Consequently, when the symmetry is spontaneously broken,  $v \neq 0$ , the pion becomes a Goldstone boson, with a vanishing mass. The  $\sigma$  particle remains massive.

In the shifted fields, the axial-vector current takes the form

$$A_\mu^i = v\partial_\mu\pi^i + (\sigma\partial_\mu\pi^i - \pi^i\partial_\mu\sigma). \quad (30.54)$$

In the tree approximation, the matrix element of the axial-vector current between the vacuum and a one-pion state is given by

$$\begin{aligned} \langle 0|A_\mu^i(x)|\pi^j(q)\rangle &= \langle 0|v\partial_\mu\pi^i(x) + \dots|\pi^j(q)\rangle = \\ &= -ip_\mu v\delta^{ij}e^{-ipx} + \dots = ip_\mu F_\pi\delta^{ij}e^{-ipx}. \end{aligned} \quad (30.55)$$

From this expression, we may immediately identify

$$F_\pi = -v + \dots. \quad (30.56)$$

Next, we consider the case when an explicit symmetry breaking parameter  $c \neq 0$ . The minimum of the potential is at

$$V'(\sigma)\Big|_{\sigma=v} = \mu^2v + \lambda v^3 - c = 0. \quad (30.57)$$

The solution of this equation is given by

$$v = \pm \sqrt{\frac{-\mu^2}{\lambda} - \frac{c}{2\mu^2}} + O(c^2) . \quad (30.58)$$

The pion mass  $M_\pi^2 = \mu^2 + \lambda v^2$  does not vanish, when  $c \neq 0$ . Instead,

$$vM_\pi^2 = -F_\pi M_\pi^2 = \mu^2 v + \lambda v^3 = c . \quad (30.59)$$

From this, we finally obtain that, in the tree approximation,

$$M_\pi^2 = -\frac{c}{F_\pi} . \quad (30.60)$$

Together with (30.49) this leads to the PCAC condition

$$\partial^\mu A_\mu^i = \frac{\pi^i}{F_\pi M_\pi^2} . \quad (30.61)$$

### 30.3.2 Pion-pion Scattering at Tree Level

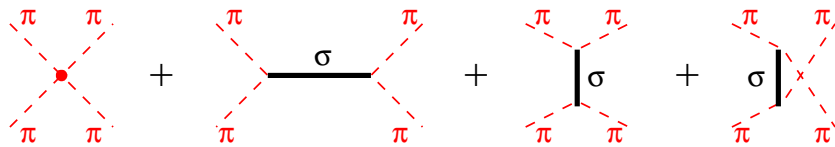
It is very instructive to consider  $\pi\pi$  scattering within the linear  $\sigma$ -model. The scattering amplitude for the process

$$\pi^i(p_1) + \pi^j(p_2) \rightarrow \pi^k(p_3) + \pi^l(p_4) , \quad (30.62)$$

owing to the Lorentz invariance and Bose-symmetry, has the following general representation

$$\begin{aligned} T^{ij,kl}(p_1, p_2, p_3, p_4) = & \delta^{ij}\delta^{kl} A(s, t, u) + \delta^{ik}\delta^{jl} A(t, u, s) + \\ & + \delta^{il}\delta^{jk} A(u, s, t) . \end{aligned} \quad (30.63)$$

Diagrams, contributing to the  $\pi\pi$  scattering amplitude in the linear  $\sigma$ -model at the tree approximation, are shown in the figure below.



After a straightforward calculation, carried out with the use of the Lagrangian displayed in (30.52), we obtain

$$A(s, t, u) = A(s) = \frac{4\lambda^2 v^2}{M_\sigma^2 - s} - 2\lambda. \quad (30.64)$$

If the symmetry breaking term  $c = 0$ , the pions are massless and the elastic threshold is located at  $s = 4M_\pi^2 = 0$ . It can be straightforwardly verified that in this case the amplitude  $A(s)$  vanishes at threshold

$$A(s) = \frac{(4\lambda^2 v^2 - 2\lambda M_\sigma^2) + 2\lambda s}{M_\sigma^2 - s} = \frac{2\lambda s}{M_\sigma^2 - s}. \quad (30.65)$$

In other words, the diagrams with the  $\sigma$  exchange and the local four-pion interaction exactly cancel at threshold. This cancellation is a direct consequence of chiral symmetry, because, due to chiral symmetry, the Goldstone bosons have *derivative interactions*, vanishing at the vanishing four-momenta. This leads to the so-called *Adler zeros* in the amplitudes with Goldstone bosons.

If  $c \neq 0$ , the amplitude takes the form

$$A(s) = \frac{2\lambda(s - M_\pi^2)}{M_\sigma^2 - s}. \quad (30.66)$$

An important remark is in order. We have seen that the pions are Goldstone bosons, whose mass is protected by chiral symmetry. The mass of the  $\sigma$  is not, and there exists no particular reason, why its mass should be small. For this reason, it is reasonable to investigate the limit  $M_\pi \ll M_\sigma$ , where the  $\sigma$ -meson decouples from physics at low energies. This makes sense from the phenomenological point of view as well, since experimentally a narrow low-lying chiral partner of pions is not observed. Performing the limit  $M_\sigma \rightarrow \infty$  in (30.66), we immediately obtain

$$A(s) = \frac{s - M_\pi^2}{F_\pi^2} + O(M_\sigma^{-4}). \quad (30.67)$$

Note that the result is *universal*: in the large- $M_\sigma$  limit, it does not contain the parameters of the linear  $\sigma$ -model anymore. Rather, the final result is expressed in terms of the physical observables  $M_\pi$  and  $F_\pi$ . As we shall see later, (30.67) is a consequence of chiral symmetry alone and holds in QCD as well.

### 30.3.3 Large $-M_\sigma$ Limit in the Linear $\sigma$ -model

It is convenient to work in the vector notations. We introduce the vector

$$\phi^A = (\sigma, \boldsymbol{\pi}) . \quad (A = 0, 1, 2, 3) \quad (30.68)$$

The Lagrangian  $\mathcal{L}_0$ , given in (30.46), in terms of the field  $\phi^A$  is rewritten in a form which is manifestly invariant with respect to the  $O(4) = SU(2) \times SU(2)$  group

$$\mathcal{L}_0 = \frac{1}{2} \partial_\mu \phi^T \partial^\mu \phi - \frac{\mu^2}{2} \phi^T \phi - \frac{\lambda}{4} (\phi^T \phi)^2 , \quad (30.69)$$

where the superscript  $T$  denotes transposed.

We proceed further by equipping this Lagrangian with external sources. Namely, we define the covariant derivative

$$\nabla_\mu \phi^A = \partial_\mu \phi^A + F_\mu^{AB} \phi^B , \quad (A, B = 0, 1, 2, 3) \quad (30.70)$$

where the external vector and axial-vector sources are related to the anti-symmetric tensor  $F_\mu^{AB}$  in the following manner

$$\begin{aligned} F_\mu^{0i} &= -F_\mu^{i0} = a_\mu^i , & (i, j = 1, 2, 3) \\ F_\mu^{ij} &= -\epsilon^{ijk} v_\mu^k . \end{aligned} \quad (30.71)$$

In addition, we introduce the four-vector  $f = (s, \mathbf{p})$ , consisting of the scalar and the pseudoscalar sources. The modified Lagrangian is given by

$$\mathcal{L} = \frac{1}{2} \nabla^\mu \phi^T \nabla_\mu \phi - \frac{\mu^2}{2} \phi^T \phi - \frac{\lambda}{4} (\phi^T \phi)^2 + f^T \phi + h \operatorname{tr} F_{\mu\nu} F^{\mu\nu} , \quad (30.72)$$

where

$$F_{\mu\nu} = [\nabla_\mu, \nabla_\nu] . \quad (30.73)$$

The last term is added to the Lagrangian, in order to ensure the finiteness of all Green's functions in the theory. Namely, if such term is not added, some Green's functions with 2, 3, 4 external vector and axial-vector legs would be divergent. Note that the constant  $h$  multiplies the operator which contains external sources only and therefore contributes a polynomial to all Green functions. Such constants are hereafter referred to as 'high-energy constants'.

This Lagrangian is invariant under the transformations

$$\begin{aligned}
\phi^0 &\mapsto \phi^0 - 2\beta^i \phi^i, \\
\phi^i &\mapsto \phi^i - 2\epsilon^{ijk} \alpha^j \phi^k + 2\beta^i \phi^0, \\
s &\mapsto s - 2\beta^i p^i, \\
p^i &\mapsto p^i - 2\epsilon^{ijk} \alpha^j p^k + 2\beta^j s, \\
a_\mu^i &\mapsto a_\mu^i + 2\partial_\mu \beta^i - 2\epsilon^{ijk} \beta^j v_\mu^k - 2\epsilon^{ijk} \alpha^j a_\mu^k, \\
v_\mu^i &\mapsto v_\mu^i + 2\partial_\mu \alpha^i + 2\epsilon^{ijk} \alpha^j v_\mu^k + 2\epsilon^{ijk} \beta^j a_\mu^k.
\end{aligned} \tag{30.74}$$

Consequently, for such a choice of the external sources, the generating functional of the linear  $\sigma$ -model obeys *exactly the same Ward identities* under  $SU_L(2) \times SU_R(2)$  transformations as the generating functional in QCD

$$Z_\sigma(s, p, v, a) = Z_\sigma(s + \delta s, p + \delta p, v + \delta v, a + \delta a). \tag{30.75}$$

Further, in analogy with the previous section, we find it useful to investigate the limit of large  $M_\sigma$ , where only the pion degrees of freedom survive in the theory. Here, we do this in tree approximation, where it suffices to substitute the solutions of the classical equations of motion into the Lagrangian. These equations take the following form

$$\nabla_\mu \nabla^\mu \phi_c = f - \mu^2 \phi_c - \lambda (\phi_c^T \phi_c) \phi_c. \tag{30.76}$$

In order to solve this equation by iterations, it is convenient to use the following parameterization of the classical field

$$\phi_c^A = -v R U^A, \quad (U^T U = 1) \tag{30.77}$$

where  $v = -\sqrt{-\mu^2/\lambda}$  and  $R$  is the 'radial' field. In this parameterization, the equations of motion can be rewritten as

$$\begin{aligned}
\Box R + R(U^T \nabla_\mu \nabla^\mu U) &= \chi_0^T U - \mu^2 R(1 - R^2), \\
R(\nabla_\mu \nabla^\mu U - U(U^T \nabla_\mu \nabla^\mu U)) &= \chi_0 - U(\chi^T U) - 2\partial_\mu R \nabla^\mu U
\end{aligned} \tag{30.78}$$

where

$$\chi_0^A = -\frac{1}{v} f^A. \tag{30.79}$$

We are going to find a solution of this equation in a form of expansion in the inverse powers of  $\mu^2$  (equivalently, of  $M_\sigma^2$ ). At lowest order, the solution is trivial:

$$R = 1. \tag{30.80}$$



At higher orders,

$$R = 1 + \delta R_1 + \delta R_2 + \cdots \quad (\delta R_n = O(\mu^{-2n})) \quad (30.81)$$

The solutions at lowest order can be easily found, solving the equations of motion iteratively

$$\begin{aligned} \delta R_1 &= -\frac{1}{2\mu^2}(\chi_0^T U + \nabla_\mu U^T \nabla^\mu U), \\ \delta R_2 &= -\frac{3}{2}\delta R_1^2 - \frac{1}{2\mu^2}\delta R_1(\nabla_\mu U^T \nabla^\mu U) + \text{total derivative}, \end{aligned} \quad (30.82)$$

and so on.

Note that the iterative procedure imposes *counting rules* on the *external sources*. Namely, the expansion is carried out in the dimensionless parameter  $p/\mu$ , where  $p$  denotes the (small) momentum of the pion. The fields  $R$  and  $U$  count both as  $O(p^0)$  and the derivative  $\partial_\mu = O(p)$ . Then, the consistent counting is achieved, if we assign

$$\begin{aligned} v_\mu &= O(p), \\ a_\mu &= O(p), \\ \chi_0 &= O(p^2). \end{aligned} \quad (30.83)$$

On the other hand, the classical action functional in terms of the fields  $R$  and  $U$  can be rewritten in the following manner

$$\begin{aligned} S(\phi_c) &= \int d^4x \mathcal{L}(\phi_c) = \\ &= -\frac{\mu^2}{2\lambda} \int d^4x \left( R(\chi_0^T U) - \frac{\mu^2}{2} R^4 + h \text{tr} F_{\mu\nu} F^{\mu\nu} \right). \end{aligned} \quad (30.84)$$

Substituting the expansion (30.81) we get

$$\begin{aligned} S(\phi_c) &= -\frac{\mu^2}{\lambda} \int d^4x \left\{ \left( \frac{1}{2} \nabla_\mu U^T \nabla^\mu U + \chi_0^T U \right) + \frac{h}{F_\pi^2} \text{tr} F_{\mu\nu} F^{\mu\nu} - \right. \\ &\quad \left. - \frac{1}{4\mu^2} (\chi_0^T U + \nabla_\mu U^T \nabla^\mu U)^2 + O(p^6) \right\}. \end{aligned} \quad (30.85)$$

Finally, expressing  $\mu^2$  through  $F_\pi$ , we obtained the desired Lagrangian up-to-and-including order  $p^4$

$$\begin{aligned} S(\phi_c) &= F_\pi^2 \int d^4x \left\{ \left( \frac{1}{2} \nabla_\mu U^T \nabla^\mu U + \chi_0^T U \right) + \frac{h}{F_\pi^2} \text{tr} F_{\mu\nu} F^{\mu\nu} + \right. \\ &\quad \left. + \frac{1}{4\lambda F_\pi^2} (\chi_0^T U + \nabla_\mu U^T \nabla^\mu U)^2 + O(p^6) \right\}. \end{aligned} \quad (30.86)$$

As anticipated, the final result contains only the pion fields, encoded in  $U$  and the external sources  $s$ ,  $p$ ,  $v$  and  $a$ . The  $\sigma$ -meson has disappeared. The effective Lagrangian contains a tower of operators at  $O(p^2)$ ,  $O(p^4)$ , etc. The coefficients of these operators are all determined through the tree-level matching in terms of the parameters  $\mu^2$ ,  $\lambda$  and  $h$  of the original model.

In general, it can be shown that there are 8 independent constants at this order. The coefficients of the operators, which contain the field  $U$ , are called *low-energy constants*, in difference with, e.g. the high-energy constant  $h$ , which multiplies the operator depending only of the external sources.

Finally, note that, due to the condition

$$U^0 U^0 + U^i U^i = 1, \quad (30.87)$$

the  $SU_L(2) \times SU_R(2)$  symmetry on the fields  $U$  is realized in the *non-linear manner* (cf. with (30.74))

$$U^i \mapsto U^i - 2\epsilon^{ijk} \alpha^j U^k + 2\beta^i \sqrt{1 - U^i U^i}. \quad (30.88)$$

The non-linear transformation law is the price to pay for having eliminated the  $\sigma$ -meson from the theory.

**Exercise 30.1:** Write the QCD Lagrangian for the three light quarks ( $u$ ,  $d$ ,  $s$ ), which is invariant under the  $U(3)_L \times U(3)_R$  transformation.

- (a) Determine corresponding to this symmetry conserved left- and right-handed currents;
- (b) In the massless case calculate the conserved vector and axial-vector current densities;
- (c) In massive quarks case calculate the divergencies of left- and right-handed as well as vector and axial-vector currents.

**Exercise 30.2:** Show that the Lagrangian of the linear  $\sigma$ -model is invariant under the isospin and axial isospin transformations.

**Exercise 30.3:** Find the new minimum for the effective  $\sigma$ -model potential when there exists a symmetry-breaking term into Lagrangian.

## Part XI

### Lecture – Beyond the SM



## Chapter 31

# Problems with SM

The SM is a mathematically-consistent renormalizable field theory which predicts or is consistent with all experimental facts. In addition:

- The SM successfully predicted the existence and form of the weak NCs;
- The SM predicted the existence and masses of the  $W$  and  $Z$  bosons;
- The SM predicted the existence of the charm quark, as necessitated by the GIM mechanism;
- The CC weak interactions, as described by the generalized Fermi theory, were successfully incorporated in the SM, as was quantum electrodynamics.
- The consistency between theory and experiment was indirectly tested by the radiative corrections and ideas of renormalization and allowed the successful prediction of the top quark mass.

Although the original formulation did not provide for massive neutrinos, they are easily incorporated by the addition of right-handed states  $\nu_R$  (Dirac) or as higher-dimensional operators, perhaps generated by an underlying seesaw (Majorana).

When combined with QCD for the strong interaction, the SM is almost certainly the approximately correct description of the elementary particles

and their interactions down to at least  $10^{-16}$  cm, with the possible exception of the Higgs sector or new very weakly coupled particles.

When combined with general relativity for classical gravity the SM accounts for most of the observed features of Nature (though not for the dark matter and energy).

However, the theory has far too much arbitrariness to be the final story. For example, the minimal version of the SM has 20 free parameters (for massless neutrinos) and another 7 (9) for massive Dirac (Majorana) neutrinos. These are 12 fermion masses (including the neutrinos), 6 mixing angles, 2  $CP$  violation phases (plus 2 possible Majorana phases), 3 gauge couplings,  $M_H$ ,  $v$ ,  $\theta_{QCD}$ ,  $M_{Pl}$ ,  $\Lambda_{Cosm}$ , minus one overall mass scale since only mass ratios are physical, not counting electric charge (i.e. hypercharge) assignments. Most physicists believe that this is just too much for the fundamental theory.

The foremost arguments in favor that the SM is not a complete theory of Nature and the existence of New Physics are that SM:

1. Has too much free parameters;
2. Does not include gravity, and therefore it cannot be valid at energy scales close the Planck scale;
3. Cannot explain the small (taking into account quantum corrections) value of the Higgs boson mass;
4. Cannot account for neutrino masses;
5. Have no good candidate for the Dark Matter and Dark Energy.

The complications of the SM can also be described in terms of a number of problems indicated in following subsections.

### 31.1 The Gauge Problem

The gauge symmetry of SM is complicated direct product of three groups,  $SU(3) \times SU(2) \times U(1)$ , with separate gauge couplings. There is no explanation for why only the EW part is chiral (parity-violating).

Similarly, the SM incorporates but does not explain another fundamental fact of nature: charge quantization, i.e. why all particles have charges which are multiples of  $e/3$ . This is important because it allows the electrical neutrality of atoms,

$$|Q_p| = |Q_e| . \quad (31.1)$$

The complicated gauge structure of the SM suggests the existence of some underlying unification of the interactions, such as one would expect in a superstring or GUT. Charge quantization can be explained in such theories, though the "wrong" values of charge emerge in some constructions due to different hypercharge embeddings or non-canonical values of  $Y$  (e.g. some string constructions lead to exotic particles with charges of  $\pm e/2$ ).

Charge quantization may also be explained, at least in part, by the existence of magnetic monopoles or the absence of anomalies (the absence of anomalies is not sufficient to determine all of the  $Y$  assignments without additional assumptions, such as family universality), but either of these is likely to find its origin in some kind of underlying unification.

## 31.2 The Fermion Problem

All matter under ordinary terrestrial conditions can be constructed out of the fermions of the first family,

$$(\nu_e, e^-, u, d) . \quad (31.2)$$

Yet we know from laboratory studies that there are more two heavier copies of the first family:

$$(\nu_\mu, \mu^-, c, s) , \quad (\nu_\tau, \tau^-, t, b) , \quad (31.3)$$

with no obvious role in Nature. The SM gives no explanation for the existence of these heavier families and no prediction for their numbers.

Furthermore, there is no explanation or prediction of the fermion masses, which are observed to occur in a hierarchical pattern which varies over 5 orders of magnitude between the  $t$  quark and the  $e^-$ , or of the quark and lepton mixings.

Even more mysterious are the neutrinos, which are many orders of magnitude lighter still. It is not even certain whether the neutrino masses are Majorana

or Dirac. A related difficulty is that, while the  $CP$  violation observed in the laboratory is well accounted for by the phase in the Kobayashi-Maskawa matrix, there is no SM source of  $CP$  breaking adequate to explain the baryon asymmetry of the universe.

There are many possible suggestions of new physics that might shed light on these questions. The existence of multiple families could be due to large representations of some string theory or GUT, or they could be associated with different possibilities for localizing particles in some higher dimensional space. The latter could also be associated with string compactifications, or by some effective brane world scenario. The hierarchies of masses and mixings could emerge from wave function overlap effects in such higher-dimensional spaces.

Another interpretation, also possible in string theories, is that the hierarchies are because some of the mass terms are generated by higher dimensional operators and therefore suppressed by powers of  $\langle 0|S|0\rangle/M_X$ , where  $S$  is some SM singlet field and  $M_X$  is some large scale such as  $M_{Pl}$ . The allowed operators could perhaps be enforced by some family symmetry. Radiative hierarchies, in which some of the masses are generated at the loop level, or some form of compositeness are other possibilities.

Despite all of these ideas there is no compelling model and none of these yields detailed predictions. Grand unification by itself doesn't help very much, except for the prediction of  $m_b$  in terms of  $m_\tau$  in the simplest versions.

The small values for the neutrino masses suggest that they are associated with Planck or grand unification physics, as in the seesaw model, but there are other possibilities.

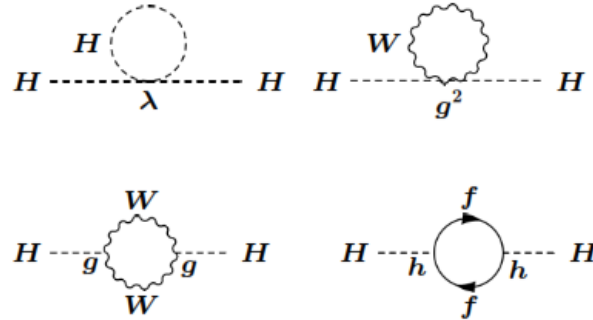
Almost any type of new physics is likely to lead to new sources of  $CP$  violation.

### 31.3 The Hierarchy Problem

In the SM one introduces an elementary Higgs field to generate masses for the  $W$ ,  $Z$ , and fermions. However, there is a complication. The tree-level (bare) Higgs mass receives quadratically-divergent corrections from the loop diagrams, including self-interactions, interactions with gauge bosons, and



interactions with fermions, see the figure below.



One finds

$$M_H^2 = (M_H^2)_{bare} + \mathcal{O}(\lambda, g^2, h^2)\Lambda^2, \quad (31.4)$$

where  $\Lambda$  is the next higher scale in the theory.

If there were no higher scale one could simply interpret  $\Lambda$  as an UV cutoff and take the view that  $M_H$  is a measured parameter, with  $(M_H)_{bare}$  not observable. However, the theory is presumably embedded in some larger theory that cuts off the momentum integral at the finite scale of the new physics.

There is no analogous fine-tuning associated with logarithmic divergences, such as those encountered in QED, because

$$\alpha \ln\left(\frac{\Lambda}{m_e}\right) < \mathcal{O}(1), \quad (31.5)$$

even for  $\Lambda = M_{Pl}$ . For example, if the next scale is gravity  $\Lambda$  is the Planck scale

$$M_{Pl} = G_N^{-1/2} \sim 10^{19} \text{ GeV}. \quad (31.6)$$

In a GUT one would expect  $\Lambda$  to be of order the unification scale

$$M_X \sim 10^{14} \text{ GeV}. \quad (31.7)$$

Hence, the natural scale for  $M_H$  is  $\mathcal{O}(\Lambda)$ , which is much larger than the expected value.

There must be a fine-tuned and apparently highly contrived cancellation between the bare value and the correction, to more than 30 decimal places

in the case of gravity. If the cutoff is provided by a GUT there is a separate hierarchy problem at the tree-level. The tree-level couplings between the Higgs field and the superheavy fields lead to the expectation that  $M_H$  is close to the unification scale unless unnatural fine-tunings are done, i.e. one does not understand why  $(M_W/M_X)^2$  is so small in the first place.

One solution to this Higgs/hierarchy problem is TeV scale SUSY, in which the quadratically-divergent contributions of fermion and boson loops cancel, leaving only much smaller effects of the order of SUSY-breaking (however, SUSY GUTs still suffer from the tree-level hierarchy problem).

There are also non-SUSY extended models in which the cancellations are between bosons or between fermions. This class includes Little Higgs models, in which the Higgs is forced to be lighter than new TeV scale dynamics because it is a pseudo-Goldstone boson of an approximate underlying global symmetry, and Twin-Higgs models.

Another possibility is to eliminate the elementary Higgs fields, replacing them with some dynamical symmetry breaking mechanism based on a new strong dynamics. In technicolor, for example, the SSB is associated with the expectation value of a fermion bilinear, analogous to the breaking of chiral symmetry in QCD. Extended technicolor, top-color, and composite Higgs models all fall into this class.

Large and/or warped extra dimensions can also resolve the difficulties, by altering the relation between  $M_{Pl}$  and a much lower fundamental scale, by providing a cutoff at the inverse of the extra dimension scale, or by using the boundary conditions in the extra dimensions to break the EW symmetry (Higgsless models). Deconstruction models, in which no extra dimensions are explicitly introduced, are closely related.

Most of the models mentioned above have the potential to generate flavor changing NC and  $CP$  violation effects much larger than observational limits. Pushing the mass scales high enough to avoid these problems may conflict with a natural solution to the hierarchy problem, i.e. one may reintroduce a little hierarchy problem. Many are also strongly constrained by precision EW physics.

In some cases the new physics does not satisfy the decoupling theorem, leading to large oblique corrections. In others new tree-level effects may again force the scale to be too high. The most successful from the precision EW

point of view are those which have a discrete symmetry which prevents vertices involving just one heavy particle, such as  $R$ -parity in SUSY,  $T$ -parity in some little Higgs models, and  $KK$ -parity in universal extra dimension models.

A very different possibility is to accept the fine-tuning, i.e. to abandon the notion of naturalness for the weak scale, perhaps motivated by anthropic considerations. This could emerge, for example, in split SUSY.

### 31.4 The Strong $CP$ Problem

Another fine-tuning problem is the strong  $CP$  problem. One can add to the QCD Lagrangian an additional term,

$$\frac{\theta_{QCD}}{32\pi^2} g_s^2 G\tilde{G} \quad (31.8)$$

which breaks  $P$ ,  $T$  and  $CP$  symmetry. Here

$$\tilde{G}_{\mu\nu}^i = \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} G^{\alpha\beta i} , \quad (31.9)$$

is the gluons dual field strength tensor. This term, if present, would induce an electric dipole moment  $d_N$  for the neutron. The rather stringent limits on the dipole moment lead to the upper bound

$$|\theta_{QCD}| < 10^{-11} . \quad (31.10)$$

The question is, therefore, why is  $\theta_{QCD}$  so small? It is not sufficient to just say that it is zero (i.e. to impose  $CP$  invariance on QCD) because of the observed violation of  $CP$  by the weak interactions.

As discussed in previous lectures, this is believed to be associated with phases in the quark mass matrices. The quark phase redefinitions which remove them lead to a shift in  $\theta_{QCD}$  by  $\mathcal{O}(10^{-3})$  because of the anomaly in the vertex coupling the associated global current to two gluons. Therefore, an apparently contrived fine-tuning is needed to cancel this correction against the bare value.

Solutions include the possibility that  $CP$  violation is not induced directly by phases in the Yukawa couplings, as is usually assumed in the SM, but is

somehow violated spontaneously.  $\theta_{QCD}$  then would be a calculable parameter induced at loop level, and it is possible to make  $\theta_{QCD}$  sufficiently small. However, such models lead to difficult phenomenological and cosmological problems. Alternately,  $\theta_{QCD}$  becomes unobservable (i.e. can be rotated away) if there is a massless  $u$  quark. However, most phenomenological estimates are not consistent with  $m_u = 0$ .

Another possibility is the Peccei-Quinn mechanism, in which an extra global  $U(1)$  symmetry is imposed on the theory in such a way that  $\theta_{QCD}$  becomes a dynamical variable which is zero at the minimum of the potential. The SSB of this symmetry, along with explicit breaking associated with the anomaly and instanton effects, leads to a very light pseudo-Goldstone boson known as an axion. Laboratory, astrophysical, and cosmological constraints suggest the range  $10^9 - 10^{12}$  GeV for the scale at which the  $U(1)$  symmetry is broken.

### 31.5 The Gravity Problem

Gravity is not fundamentally unified with the other interactions in the SM, although it is possible to graft on classical general relativity by hand. However, general relativity is not a quantum theory, and there is no obvious way to generate one within the SM context. Possible solutions include Kaluza-Klein and supergravity theories. These connect gravity with the other interactions in a more natural way, but do not yield renormalizable theories of quantum gravity. More promising are superstring theories (which may incorporate the above), which unify gravity and may yield finite theories of quantum gravity and all the other interactions.

String theories are perhaps the most likely possibility for the underlying theory of particle physics and gravity, but at present there appear to be a nearly unlimited number of possible string vacua (the landscape), with no obvious selection principle. As of this writing the particle physics community is still trying to come to grips with the landscape and its implications. Superstring theories naturally imply some form of supersymmetry, but it could be broken at a high scale and have nothing to do with the Higgs/hierarchy problem (split SUSY is a compromise, keeping some aspects at the TeV scale).

In addition to the fact that gravity is not unified and not quantized there is another difficulty, namely the cosmological constant,  $\Lambda_{\text{Cosm}}$ . The cosmological constant can be thought of as the energy of the vacuum. However,

we saw in previous lectures that the spontaneous breaking of  $SU(2) \times U(1)$  generates a value

$$\langle 0|V(v)|0\rangle = -\frac{\mu^4}{4\lambda} \quad (31.11)$$

for the expectation value of the Higgs potential at the minimum. This is a  $c$ -number which has no significance for the microscopic interactions. However, it assumes great importance when the theory is coupled to gravity, because it contributes to the cosmological constant. The cosmological constant becomes

$$\Lambda_{\text{Cosm}} = \Lambda_{\text{bare}} + \Lambda_{\text{SSB}} , \quad (31.12)$$

where

$$\Lambda_{\text{bare}} = 8\pi G_N V(0) \quad (31.13)$$

is the primordial cosmological constant, which can be thought of as the value of the energy of the vacuum in the absence of SSB.

In SM it is implicitly assumed

$$\Lambda_{\text{bare}} = 0 , \quad (31.14)$$

while  $\Lambda_{\text{SSB}}$  is the part generated by the Higgs mechanism:

$$|\Lambda_{\text{SSB}}| = 8\pi G_N |\langle 0|V|0\rangle| \sim 10^{56} \Lambda_{\text{obs}} . \quad (31.15)$$

Assuming that the dark energy is due to a cosmological constant, this value is some  $10^{56}$  times larger in magnitude than the observed one,

$$\Lambda_{\text{obs}} \sim \frac{(0.0024 \text{ eV})^4}{8\pi G_N} , \quad (31.16)$$

and it is of the wrong sign. This is clearly unacceptable. Technically, one can solve the problem by adding a constant  $+\mu^4/4\lambda$  to the Higgs potential  $V$ , so that  $V$  is equal to zero at the minimum, i.e.

$$\Lambda_{\text{bare}} = \frac{2\pi G_N \mu^4}{\lambda} . \quad (31.17)$$

However, with our current understanding there is no reason for  $\Lambda_{\text{bare}}$  and  $\Lambda_{\text{SSB}}$  to be related. The need to invoke such an incredibly fine-tuned cancellation to 50 decimal places is probably the most unsatisfactory feature of the SM.

The problem becomes even worse in superstring theories, where one expects a vacuum energy of  $\mathcal{O}(M_P^4)$  for a generic point in the landscape, leading to

$$\Lambda_{\text{obs}} \gtrsim 10^{123} |\Lambda_{\text{obs}}|. \quad (31.18)$$

The situation is almost as bad in GUTs.

So far, no compelling solution to the cosmological constant problem has emerged. One intriguing possibility invokes the anthropic (environmental) principle, i.e. that a much larger or smaller value of  $|\Lambda_{\text{Cosm}}|$  would not have allowed the possibility for life to have evolved because the Universe would have expanded or recollapsed too rapidly. This would be a rather meaningless argument unless:

- Nature somehow allows a large variety of possibilities for  $|\Lambda_{\text{Cosm}}|$  (and possibly other parameters or principles) such as in different vacua;
- There is some mechanism to try all or many of them.

In recent years it has been suggested that both of these needs may be met. There appear to be an enormous landscape of possible superstring vacua, with no obvious physical principle to choose one over the other. Something like eternal inflation could provide the means to sample them, so that only the environmentally suitable vacua lead to long-lived Universes suitable for life.

These ideas are highly controversial and are currently being heatedly debated.

## 31.6 The New Ingredients

It is now clear that the SM requires a number of new ingredients:

1. *A mechanism for small neutrino masses.* The most popular possibility is the minimal seesaw model, implying Majorana masses, but there are other plausible mechanisms for either small Dirac or Majorana masses.
2. *A mechanism for the baryon asymmetry.* The SM has neither the non-equilibrium condition nor sufficient  $CP$  violation to explain the observed asymmetry between baryons and antibaryons in the Universe.

The third necessary ingredient for creation of the baryon asymmetry, baryon number non-conservation, is present in the SM because of non-perturbative vacuum tunneling (instanton) effects. These are negligible at zero temperature where they are exponentially suppressed, but important at high temperatures due to thermal fluctuations (sphaleron configurations), before or during the EW phase transition.

One possibility involves the out of equilibrium decays of superheavy Majorana right-handed neutrinos (*leptogenesis*), as expected in the minimal seesaw model. Another involves a strongly first order EW phase transition (*EW baryogenesis*). This is not expected in the SM, but could possibly be associated with loop effects in the minimal SUSY extension (MSSM) if one of the scalar top quarks is sufficiently light. However, it is most likely in extensions of the MSSM involving SM singlet Higgs fields that can generate a dynamical  $v$  term, which can easily lead to strong first order transitions at tree-level.

Such extensions would likely yield signatures observable at the LHC. Both the seesaw models and the singlet extensions of the MSSM could also provide the needed new sources of  $CP$  violation.

Other possibilities for the baryon asymmetry include the decay of a coherent scalar field, such as a scalar quark or lepton in SUSY (the *Affleck-Dine mechanism*), or  $CPT$  violation.

Finally, one cannot totally dismiss the possibility that the asymmetry is simply due to an initial condition on the big bang. However, this possibility disappears if the universe underwent a period of *inflation*.

3. *Explanation of the dark energy.* In recent years a remarkable concordance of cosmological observations involving the cosmic microwave background radiation, acceleration of the Universe as determined by Type Ia supernova observations, large scale distribution of galaxies and clusters, and big bang nucleosynthesis has allowed precise determinations of the cosmological parameters: the Universe is close to flat, with some form of *dark energy* making up about 70% of the energy density. *Dark matter* constitutes  $\sim 25\%$ , while ordinary matter (mainly baryons) represents only about 5%.

The mysterious dark energy, which is the most important contribution to the energy density and leads to the acceleration of the expansion of the Universe, is not accounted for in the SM. It could be due to a cosmological constant that is incredibly tiny on the particle physics

scale, or to a slowly time varying field (quintessence). Is the acceleration somehow related to an earlier and much more dramatic period of inflation? If it is associated with a time-varying field, it could be connected also with a possible time variation of coupling "constants".

4. *Explanation the dark matter.* Similarly, the SM has no explanation for the observed *dark matter*, which contributes much more to the matter in the Universe than the stuff we are made of. It is likely, though not certain, that the dark matter is associated with elementary particles. An attractive possibility is weakly interacting massive particles (WIMPs), which are typically particles in the  $10^2 - 10^3$  GeV range with weak interaction strength couplings, and which lead naturally to the observed matter density.

These could be associated with the lightest SUSY partner (usually a neutralino) in SUSY models with  $R$ -parity conservation, or analogous stable particles in Little Higgs or universal extra dimension models. There are a wide variety of variations on these themes, e.g. involving very light gravitinos or other SUSY particles.

There are many searches for WIMPs going on, including direct searches for the recoil produced by scattering of Solar System WIMPs, indirect searches for WIMP annihilation products, and searches for WIMPs produced at accelerators.

Axions, perhaps associated with the strong  $CP$  problem or with string vacua, are another possibility. Searches for axions produced in the Sun, in the laboratory, or from the early universe are currently underway.

5. *Suppressions of flavor changing NCs, proton decay, and electric dipole moments.* The SM has a number of accidental symmetries and features which forbid proton decay, preserve lepton number and lepton family number (at least for vanishing neutrino masses), suppress transitions such as  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  at tree-level, and lead to highly suppressed electric dipole moments for the  $e^-$ ,  $n$ , atoms, etc.

However, most extensions of the SM have new interactions which violate such symmetries, leading to potentially serious problems with flavor changing NC and electric dipole moments. There seems to be a real conflict between attempts to deal with the Higgs/hierarchy problem and the prevention of such effects.

Recently, there has been much discussion of minimal flavor violation, which is the hypothesis that all flavor violation, even that which is as-



sociated with new physics, is proportional to the SM Yukawa matrices, leading to a significant suppression of flavor changing effects.

**Exercise 31.1:** Find the relation of the mixed angle with the mass eigenvalues in the model with  $2 \times 2$  hermitian fermion mass matrix of the form:

$$M = \begin{pmatrix} 0 & a \\ a^* & b \end{pmatrix}.$$

**Exercise 31.2:** Show that if there were a set of scalars transforming as a doublet under the weak  $SU(2)$  symmetry and as a triplet under the color  $SU(3)$ , then both the baryon number  $B$  and the lepton number  $L$  are not conserved. However, the linear combination  $B - L$  is conserved.

**Exercise 31.3:** Why in the SM it is forbidden the transition of the  $Z$  particle into two Higgs bosons.

**Exercise 31.4:** Consider a model with two left-handed leptons, two right-handed lepton doublets, two left-handed neutral leptons and usual Higgs  $SU(2) \times U(1)$  doublet. Find the expressions of weak eigenstates in terms of mass eigenstates.

**Exercise 31.5:** Compute the effective Lagrangian for the transition

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

in the modified SM with massive neutrinos.

**Exercise 31.6:** Neutrons are neutral particle just as neutral  $K$ -mesons. Why is it not meaningful to introduce linear combinations of  $n_1$  and  $n_2$ , similar to the  $K_1^0$  and  $K_2^0$ ?

**Exercise 31.7:** If the baryon number is conserved, the neutron oscillation,  $n \rightarrow \bar{n}$ , is forbidden. The experimental limit on the time scale of such oscillations in free space and zero magnetic field is

$$\tau_{n-\bar{n}} \geq 3 \times 10^6 \text{ s}.$$

Let  $H_0$  be the Hamiltonian of the world in the absence of any interaction which mixes  $n$  and  $\bar{n}$ . Then

$$H_0|n\rangle = m_n|n\rangle, \quad H_0|\bar{n}\rangle = m_n|\bar{n}\rangle$$

for states at rest. Let  $H'$  be the interaction which turns  $n$  into  $\bar{n}$  and vice versa:

$$H'|n\rangle = \epsilon|\bar{n}\rangle, \quad H'|\bar{n}\rangle = \epsilon|n\rangle,$$

where  $\epsilon$  is real and  $H'$  does not flip spin.

(a) Start with a neutron at  $t = 0$  and calculate the probability that it will be observed to be an antineutron at time  $t$ . When the probability is first equal to 50%, call that time  $\tau_{n-\bar{n}}$ . In this way convert the experimental limit on  $\tau_{n-\bar{n}}$  into a limit on  $\epsilon$ ;

(b) Reconsider the problem in the presence of the earth's magnetic field

$$B = 0.5 \text{ Gs}.$$

The magnetic moment of the neutron is

$$\mu_n \approx -6 \times 10^{-18} \text{ MeV/Gs}.$$

The magnetic moment of the antineutron is opposite. Begin with a neutron at  $t = 0$  and calculate the probability it will be observed to be an antineutron at time  $t$ . Ignore possible radioactive transitions.

## Chapter 32

# Flavor and CP Violation in EFT

As it was mentioned in the first chapter of this lecture, several observations suggest to replace the SM by a new theory already around the TeV scale. Given that non-renormalizable terms must be added to the SM Lagrangian, i.e. the SM probably is only an effective low energy theory.

If we assume that the new degrees of freedom, which complete the theory of Nature, are heavier than the SM particles, we can integrate them out and describe physics beyond the SM by means of an *Effective Field Theory* (EFT) approach. Then the SM Lagrangian becomes the renormalizable part of this generalized Lagrangian which includes an infinite sum of operators with dimension  $d \geq 5$ , constructed in terms of SM fields and suppressed by inverse powers of the new physics scale  $\Lambda \gg v$ . This approach is a generalization of the Fermi theory of weak interactions, where the dimension six four-fermion operators describing weak decays are the results of having integrated out the  $W$  boson.

The generic Lagrangian in EFT approach to the SM reads

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_n^{\substack{d \geq 5 \\ n}} \frac{c_n^d}{\Lambda^{d-4}} \mathcal{O}_n^{(d)}(\text{SM}), \quad (32.1)$$

where  $\mathcal{O}_n^{(d)}(\text{SM})$  are operators of dimension  $d \geq 5$  containing SM fields only and compatible with the SM gauge symmetry.

### 32.1 The Flavor Sector of the SM

Flavor physics is the study of different generations, or 'flavors', of quarks and leptons, their spectrum and their transitions. In SM there are six different types of quarks: up ( $u$ ), down ( $d$ ), strange ( $s$ ), charm ( $c$ ), bottom ( $b$ ) and top ( $t$ ) and three different type of charged leptons: electron, muon and tau. We know that in the SM

$$m_t \gg m_c \gg m_u, \quad m_b \gg m_s \gg m_d, \quad (32.2)$$

and that the same hierarchies hold for the Higgs Yukawa couplings with quarks and leptons. Our lack of understanding of why nature has exactly three generations of quarks and leptons and why their properties (masses and mixing angles) are described by such hierarchical values is the so called *SM flavor puzzle*. In the limit of unbroken EW symmetry none of the basic constituent of matter would have a non-zero mass. The SM flavor puzzle is, therefore, intimately related to the other big open question in particle physics, i.e. what is the exact mechanism behind SSB.

Once the SM quark and lepton masses (as well as quark mixing angles, plus a phase) have been fixed, the SM is a highly predictive theory for flavor transitions. Particularly, any flavor transition has to involve the exchange of at least a  $W$  boson and therefore flavor changing neutral transitions can only arise (at least) at the loop-level.

In the last few years tremendous progress has been reached in testing the mechanism of quark flavor mixing by several experiments, LHCb and  $B$ -factories (Belle and Babar) as well as the high energy experiments (ATLAS and CMS), finding good agreement with the SM expectations. At the same time, there are a few flavor measurements that could be interpreted as tantalizing hints for deviations if compared to the SM predictions. Particularly, there have been a lot of attention on the anomalies in angular observables in the decay  $B_d \rightarrow K^* \mu \mu$  (involving a  $b \rightarrow s$  flavor transition), as observed by the LHCb collaboration, as well as on the observables testing lepton flavor universality,  $\text{BR}(B \rightarrow K \mu \mu) / \text{BR}(B \rightarrow K e e)$ , as observed at LHCb and on the rare decays  $B \rightarrow D \tau \nu_\tau$  and  $B \rightarrow D^* \tau \nu_\tau$  by Belle, Babar, and LHCb.

Present and future flavor measurements will be able to probe new physics. Observing new sources of flavor mixing is, in fact, a natural expectation for any extension of the SM with new degrees of freedom not far from the TeV scale. While direct searches of new particles at high energies provide

information on the mass spectrum of the possible new degrees of freedom, the indirect information from low energy flavor observables translates into unique constraints on their couplings.

The SM Lagrangian can be divided in three main parts: the gauge, the Higgs, and the flavor sector. The first two parts are highly symmetric,

$$\begin{aligned} \mathcal{L}_{\text{gauge+Higgs}} = & i \sum_{\psi} \bar{\psi}^i \gamma^\nu D_\nu \psi_i - \frac{1}{4} G_{\mu\nu}^a G^{\mu\nu}_a - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \\ & + |D_\mu \phi|^2 + (\mu^2 |\phi|^2 - \lambda |\phi|^4), \end{aligned} \quad (32.3)$$

and fully determined by a small set of free parameters: the three gauge couplings,  $g_1$ ,  $g_2$  and  $g_s$ , corresponding to the SM gauge groups  $SU(3) \times SU(2) \times U(1)_Y$ , the Higgs mass and the vacuum expectation value  $v$  (or, equivalently, the Higgs mass term,  $\mu$ , and the quartic coupling,  $\lambda$ ). In (32.3)  $G$ ,  $W$ , and  $B$  are the  $SU(3)$ ,  $SU(2)$ , and  $U(1)_Y$  gauge fields, respectively, and we have defined the quark and lepton field content as

$$\psi^i \equiv Q_L^i, \quad L_L^i, \quad u_R^i, \quad d_R^i, \quad e_R^i, \quad (32.4)$$

with

$$\begin{aligned} Q_L^i = (3, 2, \frac{1}{6}), \quad L_L^i = (1, 2, -\frac{1}{2}), \\ u_R^i = (3, 1, \frac{2}{3}), \quad d_R^i = (3, 1, -\frac{1}{3}), \quad e_R^i = (1, 1, -1), \end{aligned} \quad (32.5)$$

where  $i = 1, 2, 3$  is the flavor (or generation) index and the three numbers refer to the representation under the SM gauge group.

The Lagrangian (32.3) possesses a large flavor symmetry that can be decomposed as

$$\begin{aligned} SU(3)^5 \times U(1)^5 = & SU(3)_q^3 \times SU(3)_l^2 \times \\ & \times U(1)_B \times U(1)_L \times U(1)_Y \times \\ & \times U(1)_{\text{QCD}} \times U(1)_e, \end{aligned} \quad (32.6)$$

which contains  $SU(3)$  quark and lepton subgroups. The three  $U(1)$  symmetries (labeled by  $B$ ,  $L$  and  $Y$ ) can be identified with baryon and lepton numbers and hypercharge, the latter of which is broken spontaneously by the Higgs field. The two remaining  $U(1)$  groups (labeled by QCD and  $e$ ) can be identified with the Peccei-Quinn symmetry and with a global rotation of a single  $SU(2)$  singlet  $e_R$ .

The flavor sector of the SM Lagrangian breaks the  $SU(3)^5$  symmetry through the Yukawa interactions

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^d \bar{Q}_L^i \phi D_R^j - Y_{ij}^u \bar{Q}_L^i \tilde{\phi} U_R^j - Y_{ij}^e \bar{L}_L^i \phi e_R^j + \text{h.c.} , \quad (32.7)$$

where  $\phi = (1, 2, 1/2)$  is the Higgs field,  $\tilde{\phi}$  is its conjugate representation ( $\tilde{\phi} = i\sigma_2 \phi^\dagger$ ) and  $Y^{d,u,e}$  are the three Yukawa couplings.

The diagonalization of each Yukawa coupling requires a bi-unitary transformation. Particularly, in the absence of right-handed neutrinos the lepton sector Yukawa can be fully diagonalized by the transformation

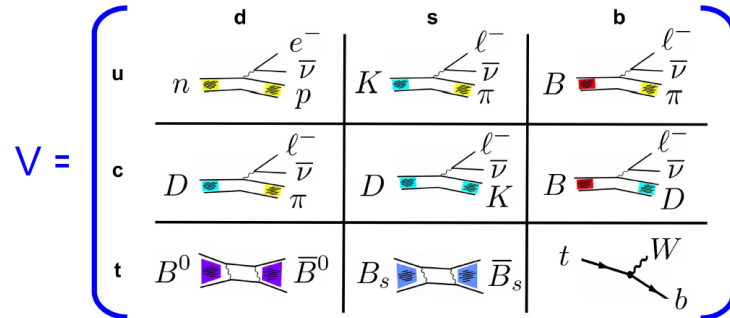
$$U_{eL} Y_e U_{eR}^\dagger = \text{diag}(y_e^1, y_e^2, y_e^3) = \frac{\sqrt{2}}{v} \text{diag}(m_e, m_\mu, m_\tau) . \quad (32.8)$$

In the quark sector, it is not possible to simultaneously diagonalize the two Yukawa matrices  $Y_u$  and  $Y_d$  without breaking the  $SU(2)$  gauge invariance. If, for example, we choose the basis in which the up Yukawa is diagonal, then

$$\begin{aligned} Y_u &= \text{diag}(y_u^1, y_u^2, y_u^3) = \frac{\sqrt{2}}{v} (m_u, m_c, m_t) , \\ Y_d &= V \cdot \text{diag}(y_d^1, y_d^2, y_d^3) = \frac{\sqrt{2}}{v} V \cdot (m_d, m_s, m_b) , \end{aligned} \quad (32.9)$$

where  $V = U_{uL} U_{dL}^\dagger$  denotes the Kobayashi-Maskawa matrix.

On the figure below you can see the list of the most sensitive observables used to determine the several elements of the Kobayashi-Maskawa  $3 \times 3$  matrix.



In the SM the  $SU(2)$  gauge symmetry is broken spontaneously by the Higgs field and therefore, we can equivalently rotate both left-handed up and

down quarks independently, diagonalizing simultaneously up and down quark masses. By performing these transformations, the Kobayashi-Maskawa dependence moves into the couplings of up and down quarks with the  $W$  boson. In particular, the CC part of the quark covariant derivative in (32.3) can be rewritten in the mass eigenstate basis as

$$-\frac{g}{2}\bar{Q}_L^i\gamma^\mu W_\mu^a\tau^a Q_L^i \xrightarrow{\text{mass-basis}} -\frac{g}{\sqrt{2}}(\bar{u}_L \ \bar{c}_L \ \bar{t}_L)\gamma^\mu W_\mu^+ V \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}. \quad (32.10)$$

This equation shows that the appearance of  $W$  boson flavor changing couplings, which is the only flavor changing interaction in the SM.

We can therefore conclude that, in the SM:

1. The only interactions mediating flavor changing transitions are the charged interactions;
2. There are no tree-level flavor changing neutral interactions.

In spite of the first point here, it must be stressed that the Kobayashi-Maskawa matrix,  $V$ , originates from the Yukawa sector and in absence of Yukawa couplings,  $V_{ij} = \delta_{ij}$ , therefore we have no flavor changing transitions.

### 32.1.1 Meson Mixing and the GIM Mechanism

In the SM, in order for a flavor transition to take place, the exchange of at least a virtual  $W$  is necessary. From the other hand, *flavor changing NC* process is a process in which the electric charge does not change between initial and final states. As a consequence, in the SM such processes have a reduced rate relative to a normal weak interaction process. Flavor changing NC are, however, not only suppressed by the loop, but also by the so called Glashow-Iliopoulos-Maiani (GIM) mechanism. We will explain this mechanism through the discussion of meson mixing.

Let us take the  $K$  ( $d\bar{s}$ ) and  $\bar{K}$  ( $\bar{d}s$ ) meson system. These two flavor eigenstates are not mass eigenstates and, therefore, they mix. The leading order contributions to the mixing arise from box diagrams mediated by the exchange of the  $W$  boson and the up quarks. The corresponding effective

Hamiltonian responsible of this mixing is given by

$$\frac{G_F^2 m_W^2}{16\pi^2} \left[ \sum_i F(x_i, x_i) \lambda_i^2 + \sum_{i \neq j} F(x_i, x_j) \lambda_i \lambda_j \right] [\bar{s} \gamma_\mu (1 - \gamma_5) d]^2 . \quad (32.11)$$

Here  $i$  and  $j$  runs over  $u, c$  and  $t$ ,  $F(x_i, x_j)$  are loop functions,  $x_q \equiv m_q^2/m_W^2$  and  $\lambda_i = V_{is}^* V_{id}$ . In the limit of exact flavor symmetry ( $m_d = m_s = m_b$ ) the several diagrams cancel, thanks to the unitarity of the Kobayashi-Maskawa matrix. This is the so called *GIM mechanism*, that can be applied not only to the Kaon mixing system but to all SM flavor transitions.

Historically, in 1970, at the time the GIM mechanism was proposed, only three quarks (up, down, and strange) were thought to exist. The GIM mechanism however, required the existence of a fourth quark, the charm, to explain the large suppression of flavor changing NC processes.

The breaking of the flavor symmetry induces a mass difference between the quarks, so the sum of the diagrams responsible for meson mixing will be non-zero. One can use the unitarity relations to eliminate the terms in the effective Hamiltonian that depend on  $\lambda_u$ , obtaining

$$\frac{G_F^2 m_W^2}{16\pi^2} [S_0(x_t) \lambda_t^2 + S_0(x_c) \lambda_c^2 + 2S_0(x_c, x_t) \lambda_c \lambda_t] [\bar{s} \gamma_\mu (1 - \gamma_5) d]^2 , \quad (32.12)$$

with  $S_0(x_i)$  and  $S_0(x_i, x_j)$  given by the combinations

$$\begin{aligned} S_0(x_i) &\equiv F(x_i, x_i) + F(x_u, x_u) - 2F(x_i, x_u) , \\ S_0(x_i, x_j) &\equiv F(x_i, x_j) + F(x_u, x_u) - F(x_i, x_u) - F(x_j, x_u) . \end{aligned} \quad (32.13)$$

All terms of the effective Hamiltonian (32.12) are suppressed not only the loop factor, but also the small Kobayashi-Maskawa elements, particularly suppressing the top loop contribution, and the small mass ratio  $m_c^2/m_W^2$  in the case of the charm loop contribution, as predicted by the GIM mechanism.

## 32.2 CP Violation in Meson Decays

The time evolution of the Kaon anti-Kaon system,  $\psi = (K, \bar{K})$ , can be described by the equation

$$i \frac{d\psi(t)}{dt} = \hat{H} \psi(t) , \quad (32.14)$$



where the effective Hamiltonian reads

$$\hat{H} = \hat{M} - i\frac{\hat{\Gamma}}{2} = \begin{pmatrix} M - i\Gamma/2 & M_{12} - i\Gamma_{12}/2 \\ M_{12}^* - i\Gamma_{12}^*/2 & M - i\Gamma/2 \end{pmatrix}, \quad (32.15)$$

with  $M$  and  $\Gamma$  the average mass and width of the two Kaons, respectively. The two eigenstates of the system (heavy and light, or, equivalently, long and short) have a mass and width given by

$$M_{H,L} = M \pm \text{Re}(Q), \quad \Gamma_{H,L} = \Gamma \mp 2\text{Im}(Q),$$

where

$$Q = \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right) \left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (32.16)$$

and are a linear combination of the two  $K$  and  $\bar{K}$  states

$$|K_{H,L}\rangle = p|K\rangle \mp q|\bar{K}\rangle, \quad \frac{q}{p} = -\frac{2M_{12}^* - i\Gamma_{12}^*}{2\text{Re}(Q) + 2i\text{Im}(Q)}. \quad (32.17)$$

The difference in mass of the two Kaon states,  $\Delta M_K$ , can be computed from the effective Hamiltonian in (32.12) by

$$m_K \Delta M_K = 2m_K \text{Re}(M_{12}) = \text{Re}(\langle \bar{K} | \mathcal{H}_K | K \rangle), \quad (32.18)$$

with  $m_K$  the average Kaon mass. Lattice QCD is essential to compute the matrix element of the four quark operator calculated between two quark bound states. We have

$$\langle \bar{K} | (\bar{s}\gamma_\mu(1 - \gamma_5)d)^2 | K \rangle = \frac{8}{3} B_K(\mu) F_K^2 m_K^2, \quad (32.19)$$

with  $F_K$  the Kaon decay constant and  $B_K(\mu)$  the Kaon bag parameter, evaluated at the scale  $\mu$ . Putting these pieces together and including QCD corrections, one can find

$$M_{12} = \frac{G_F^2}{12\pi^2} F_K^2 \hat{B}_K m_K m_W^2 [(\lambda_c^*)^2 \eta_1 S_0(x_c) + (\lambda_t^*)^2 \eta_2 S_0(x_t) + 2\lambda_c^* \lambda_t^* \eta_3 S_0(x_c, x_t)], \quad (32.20)$$

where  $\eta_{1,2,3}$  are QCD correction factors and we have defined the renormalization group invariant parameter

$$\hat{B}_K = \frac{B_K(\mu)}{\alpha_s(\mu)^{2/9}} \left[ 1 + \frac{\alpha_s(\mu)}{4\pi} J_3 \right], \quad (32.21)$$

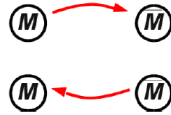
with  $J_3 \sim 1.9$ .

All  $CP$ -violating observables in  $K$  and  $\bar{K}$  (and in any meson system,  $M - \bar{M}$ ) decays to final states  $f$  and  $\bar{f}$  can be expressed in terms of phase-convention-independent combinations of  $A_f$ ,  $\bar{A}_f$ ,  $A_{\bar{f}}$  and  $\bar{A}_{\bar{f}}$ , together with  $q/p$  of (32.17), in the case of neutral-mesons, where we define

$$\begin{aligned} A_f &= \langle f | \mathcal{H} | M \rangle, & \bar{A}_f &= \langle f | \mathcal{H} | \bar{M} \rangle, \\ A_{\bar{f}} &= \langle \bar{f} | \mathcal{H} | M \rangle, & \bar{A}_{\bar{f}} &= \langle \bar{f} | \mathcal{H} | \bar{M} \rangle. \end{aligned} \quad (32.22)$$

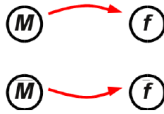
We distinguish three types of  $CP$ -violating effects in a meson  $M$  decays to a final state  $f$ :

(a)  $CP$  violation in mixing:



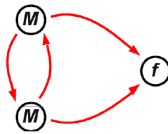
Defined by  $|q/p| \neq 1$  and arising when the two neutral mass eigenstate admixtures cannot be chosen to be  $CP$ -eigenstates;

(b)  $CP$  violation in the decay of mesons:



Defined by  $|\bar{A}_{\bar{f}}/A_f| \neq 1$ ;

(c)  $CP$  violation in interference between a decay without mixing,  $M \rightarrow f$ , and a decay with mixing  $M \rightarrow \bar{M} \rightarrow f$ :



This is defined by  $\text{Im}(q\bar{A}_f/pA_f) \neq 0$ .

One example of  $CP$  violation in mixing (a) is the asymmetry in charged-current semi-leptonic neutral meson decays for which the 'wrong sign' decays (i.e. decays to a lepton of charge opposite to the sign of the charge of the original  $b$  quark) are allowed only if there is a mixing between the meson and the anti-meson. For example, for a  $B_d$  meson

$$a_{\text{SL}}^d = \frac{\Gamma(\bar{B}_d(t) \rightarrow \ell^+ \nu X) - \Gamma(\bar{B}_d(t) \rightarrow \ell^- \bar{\nu} X)}{\Gamma(\bar{B}_d(t) \rightarrow \ell^+ \nu X) + \Gamma(\bar{B}_d(t) \rightarrow \ell^- \bar{\nu} X)} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \quad (32.23)$$

D0 collaboration performed several measurements of these asymmetries in  $B$  decays. Combining all measurements, there is a long-standing anomaly with the SM prediction in the  $a_{\text{SL}}^s - a_{\text{SL}}^d$  plane with a significance at the level of  $\sim 2 - 3\sigma$ , mainly arising from the D0 measurement of the like-sign dimuon charge asymmetry.

In charged meson decays, where mixing effects are absent, the  $CP$  violation in decay (b) is the only possible source of  $CP$  asymmetries. For example, in the  $B$  meson system:

$$a_{f^\pm} = \frac{\Gamma(B^+ \rightarrow f^+) - \Gamma(B^- \rightarrow f^-)}{\Gamma(B^+ \rightarrow f^+) + \Gamma(B^- \rightarrow f^-)} = \frac{1 - |\bar{A}_{f^-}/A_{f^+}|^2}{1 + |\bar{A}_{f^-}/A_{f^+}|^2}. \quad (32.24)$$

These asymmetries are different from zero only if at least two terms of the amplitude have different weak phases and different strong phases. Strong phases do not violate  $CP$ . Their origin is the contribution from intermediate on-shell states in the decay process that is an absorptive part of an amplitude. Non-zero  $CP$  asymmetries have been observed in a few  $B$  meson decay modes by the LHCb collaboration:  $B^+ \rightarrow K^+ K^- K^+$  and  $B^+ \rightarrow K^+ K^- \pi^+$ .

$CP$  violation in interference (c) is measured through the decays of neutral mesons and anti-mesons to a final state that is a  $CP$  eigenstate ( $f_{CP}$ )

$$\begin{aligned} a_{f_{CP}} &= \frac{\Gamma(\bar{M}(t) \rightarrow f_{CP}) - \Gamma(M(t) \rightarrow f_{CP})}{\Gamma(\bar{M}(t) \rightarrow f_{CP}) + \Gamma(M(t) \rightarrow f_{CP})} \simeq - \\ &\simeq \text{Im}(\lambda_{CP}) \sin(\Delta M_M t), \end{aligned} \quad (32.25)$$

where we have defined

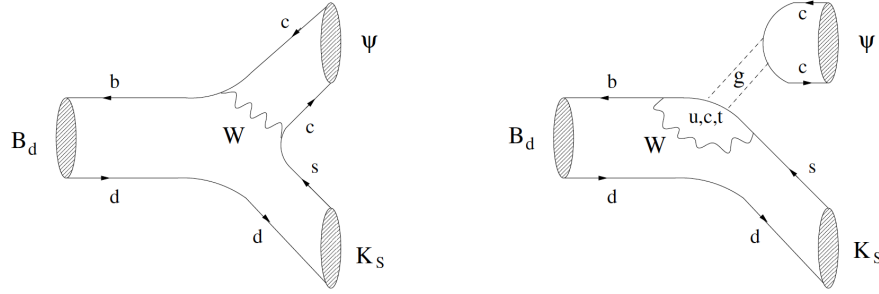
$$\lambda_{CP} = \frac{q\bar{A}_f}{pA_f} \quad (32.26)$$

and  $\Delta M_M$  is the difference in mass of the meson anti-meson system. This type of  $CP$  violation has been observed in several  $B$  meson decays, as for

example in  $B_d \rightarrow J/\psi K_S$  at Babar, Belle and by now by the LHCb, as well, leading to the measurement of the  $\beta$  angle of the Kobayashi-Maskawa matrix

$$a_{J/\psi K_S} \simeq \sin(2\beta) \sin(\Delta M_d t) . \quad (32.27)$$

The Feynman diagrams contributing to this asymmetry are given in the figure below, where we show the tree (left panel) and the penguin (right panel) contributions to  $B_d \rightarrow \Psi K_S$ .



The current world average on the angle  $\beta$  is

$$\sin(2\beta) = 0.69 \pm 0.02 . \quad (32.28)$$

The corresponding  $CP$  asymmetry in  $B_s$  decay is  $B_s \rightarrow \psi\phi$ . The SM prediction is suppressed compared to the  $\beta$  angle by  $\lambda^2$ , leading to

$$\beta_s^{\text{SM}} = 0.01882 \pm 0.0004 . \quad (32.29)$$

A summary of all mixing angle measurements in the  $B_s - \bar{B}_s$  system is

$$\beta_s = -\frac{1}{2}\phi_s^{c\bar{c}s} = -0.00165 \pm 0.00165 . \quad (32.30)$$

### 32.3 Flavor Transitions

An essential feature of flavor physics is its capability to probe very high scales, beyond the kinematical reach of high energy colliders. At the same time, flavor physics can teach us about properties of TeV-scale new physics

(i.e. how new particles couple to the SM degrees of freedom), offering complementarity with searches of new physics at colliders.

As we mentioned above, flavor changing neutral processes in the SM are highly suppressed, both because they arise at least at the loop-level and because of the GIM mechanism that introduces the dependence of these processes on the small Kobayashi-Maskawa off-diagonal elements and on the small quark masses. Then flavor transitions offer a unique opportunity to test the new physics flavor structure. Generically beyond SM models predict too large contributions to flavor transitions (the new physics flavor problem) leading us to conclude that, if TeV-scale new physics exists, it must have a highly non generic flavor structure, as for example it can obey to the *Minimal Flavor Violation* (MFV) principle.

Let us mention some experiments that are running, and will be running in the coming years, to explore flavor violation transitions. Some of the golden channels are:

- More precise measurement of the clean rare decays  $B_s \rightarrow \mu^+ \mu^-$  and  $B_d \rightarrow \mu^+ \mu^-$  at LHCb, ATLAS and CMS. The ratio of branching ratios will give us more insights on the validity of the MFV ansatz;
- Additional tests of the lepton universality relations in  $B$  decays at LHCb and Belle II:

$$\frac{\Gamma(B \rightarrow Jee)}{\Gamma(B \rightarrow J\mu\mu)} \cdot \quad (J = K, K^*, X_s, K\pi, \dots) \quad (32.31)$$

These are particularly clean tests of the SM, as the theory predictions are known to a very good precision and are not affected by hadronic uncertainties;

- Better measurements of  $B \rightarrow D\tau\nu$  and  $B \rightarrow D^*\tau\nu$ , to confirm or disprove the present anomaly in these decays, as observed at Belle, Babar and LHCb;
- Measurements of  $B \rightarrow K^{(*)}\nu\nu$  and  $K \rightarrow \pi\nu\nu$  at Belle II and KOTO, respectively;
- Additional searches of top and Higgs flavor violating couplings at LHC.

These channels (and several others) will be able either to set interesting constraints on new physics, or to shed light into the existence of new degrees of freedom beyond the SM.

### 32.3.1 The New Physics Flavor Puzzle

Let us discuss the flavor constraints on a new physics scale,

$$\Lambda \gg v , \quad (32.32)$$

associated to the higher dimensional operators contributing to flavor transitions.

Generically, one would expect the in (32.1) the Wilson coefficients

$$c_n^d = \mathcal{O}(1) , \quad (32.33)$$

however several of these operators contribute to flavor-changing processes and should be very suppressed to be in agreement with low energy flavor experiments. This is often denoted as the *new physics flavor puzzle*.

As an example, we consider the dimension 6 operators contributing to Kaon mixing:

$$\begin{aligned} O_1^{\text{VLL}} &= (\bar{s}\gamma_\mu P_L d)^2 , \\ O_1^{\text{LR}} &= (\bar{s}\gamma_\mu P_L d)(\bar{s}\gamma^\mu P_R d) , \\ O_2^{\text{LR}} &= (\bar{s}P_L d)(\bar{s}P_R d) , \\ O_1^{\text{SLL}} &= (\bar{s}P_L d)^2 , \\ O_2^{\text{SLL}} &= (\bar{s}\gamma_{\mu\nu} P_L d)(\bar{s}\gamma^{\mu\nu} P_L d) , \end{aligned} \quad (32.34)$$

plus the corresponding ones with the exchange

$$P_L \rightarrow P_R . \quad \left( P_{L,R} = \frac{1}{2}(1 \mp \gamma_5) \right) \quad (32.35)$$

The only operator that arises in the SM is  $O_1^{\text{VLL}}$ .

As an example, a toy model containing a TeV scale new  $Z'$  gauge boson with coupling

$$g' Z'_\mu [\bar{s}\gamma^\mu (1 - \gamma_5) d] \quad (32.36)$$

would produce a contribution to the operator  $O_1^{\text{VLL}}$  and, therefore, to the difference in mass of Kaon and anti-Kaon system that is equal to

$$\Delta M_K = \Delta M_K^{\text{SM}} + \frac{8}{3} m_K F_K^2 \hat{B}_K \frac{(g')^2}{m_{Z'}^2} , \quad (32.37)$$

where  $\Delta M_K^{\text{SM}}$  is the value predicted by the SM. For TeV-scale  $Z'$ s coupled to a bottom and a strange quark with a EW strength coupling, the second piece of this equation is  $\sim 4$  orders of magnitude larger than the SM contribution, and therefore, such gauge bosons are completely ruled out by Kaon mixing measurements.

This shows the tension between a generic new physics at around the TeV scale with EW-strength flavor violating couplings and low energy flavor measurements (new physics flavor puzzle).

A summary of the bounds for the four neutral meson systems ( $K$ ,  $B_d$ ,  $B_s$  and  $D$ ) is shown in the table below.

Operator	Bounds on $\Lambda$ in TeV ( $c_n^6 = 1$ )		Bounds on $c_n^6$ ( $\Lambda = 1$ TeV)	
	Re	Im	Re	Im
$(\bar{s}_L \gamma^\mu d_L)^2$	$9.8 \times 10^2$	$1.6 \times 10^4$	$9.0 \times 10^{-7}$	$3.4 \times 10^{-9}$
$(\bar{s}_R d_L)(\bar{s}_L d_R)$	$1.8 \times 10^4$	$3.2 \times 10^5$	$6.9 \times 10^{-9}$	$2.6 \times 10^{-11}$
$(\bar{c}_L \gamma^\mu u_L)^2$	$1.2 \times 10^3$	$2.9 \times 10^3$	$5.6 \times 10^{-7}$	$1.0 \times 10^{-7}$
$(\bar{c}_R u_L)(\bar{c}_L u_R)$	$6.2 \times 10^3$	$1.5 \times 10^4$	$5.7 \times 10^{-8}$	$1.1 \times 10^{-8}$
$(\bar{b}_L \gamma^\mu d_L)^2$	$6.6 \times 10^2$	$9.3 \times 10^2$	$2.3 \times 10^{-6}$	$1.1 \times 10^{-6}$
$(\bar{b}_R d_L)(\bar{b}_L d_R)$	$2.5 \times 10^3$	$3.6 \times 10^3$	$3.9 \times 10^{-7}$	$1.9 \times 10^{-7}$
$(\bar{b}_L \gamma^\mu s_L)^2$	$1.4 \times 10^2$	$2.5 \times 10^2$	$5.0 \times 10^{-5}$	$1.7 \times 10^{-5}$
$(\bar{b}_R s_L)(\bar{b}_L s_R)$	$4.8 \times 10^2$	$8.3 \times 10^2$	$8.8 \times 10^{-6}$	$2.9 \times 10^{-6}$

These bounds, assuming an effective coupling  $c_n^6/\Lambda^2$ , quoted for  $\Lambda$  are obtained setting  $|c_n^6| = 1$ ; those for  $c_{\text{New Phys.}}$  are obtained setting  $\Lambda = 1$  TeV. We define

$$q_{L,R} \equiv P_{L,R} q . \quad (32.38)$$

In the first two entries the bounds on the new physics scale,  $\Lambda$ , having fixed the absolute value of the corresponding Wilson coefficient ( $c_n^6$  of (32.1)) to one (the first column is for  $c_n^6 = 1$ , the second one for  $c_n^6 = i$ ); the last two columns represent, instead, the bound on real part and on the imaginary part of the Wilson coefficient, fixing the new physics scale to 1 TeV.

A few comments are in order. The bounds are weakest (strongest) for  $B_s$  ( $K$ ) mesons, as mixing is the least (most) suppressed in the SM in that case. The

bounds on the operators with a different chirality (left-right or right-left) are stronger, especially in the Kaon case, because of the larger hadronic matrix elements. Throughout the table, bounds on the new physics scale  $\Lambda$  exceed the TeV scale by several orders of magnitude. Therefore, we can conclude that, if new physics exists at around the TeV scale, it has to possess a highly non-generic flavor structure, to explain  $c_n^d \ll 1$ .

### 32.3.2 The Minimal Flavor Violation Ansatz

TeV scale new physics could be invariant under some flavor symmetry, more easily in agreement with low energy flavor measurements. One example of a class of such models are theories with *Minimal Flavor Violation* (MFV). Under this assumption, flavor violating interactions are linked to the known structure of the SM Yukawa couplings also beyond the SM.

More specifically, the MFV ansatz can be implemented within the generic effective Lagrangian (32.1), as well as to UV complete models, and it consists of two ingredients: a flavor symmetry and a set of symmetry-breaking terms.

The symmetry is the SM global symmetry in absence of Yukawa couplings. Since this global symmetry, and particularly the  $SU(3)$  subgroups controlling quark flavor-changing transitions, is broken within the SM, it cannot be promoted to an exact symmetry of the new physics model. Particularly, in the SM we can formally recover the flavor invariance under  $\mathcal{G}_{\text{flavor}}$  by promoting the Yukawa couplings  $Y_d$ ,  $Y_u$  and  $Y_e$  to dimensionless auxiliary fields (spurions) transforming under

$$SU(3)_q^3 = SU(3)_Q \times SU(3)_U \times SU(3)_D \quad (32.39)$$

and under

$$SU(3)_\ell^2 = SU(3)_L \times SU(3)_e \quad (32.40)$$

as

$$Y_Q \sim (3, 1, \bar{3})_{SU(3)_q^3}, \quad Y_u \sim (3, \bar{3}, 1)_{SU(3)_q^3}, \quad Y_e \sim (3, \bar{3})_{SU(3)_\ell^2}. \quad (32.41)$$

Employing an effective field theory language, a theory satisfies the MFV ansatz, if all higher-dimensional operators, constructed from SM and  $Y_{u,d,e}$  fields, are invariant under the flavor group,  $\mathcal{G}_{\text{flavor}}$ .



The invariance under  $CP$  of the new physics operators may or may not be imposed in addition to this criterion. In the down quark sector, the several operators will be combinations of the invariants

$$\bar{Q}_L Y_u Y_u^\dagger Q_L, \quad \bar{D}_R Y_d^\dagger Y_u Y_u^\dagger Q_L, \quad \bar{D}_R Y_d^\dagger Y_u Y_u^\dagger Y_d D_R. \quad (32.42)$$

As an example, let us take the operators in (32.34) and impose the MFV hypothesis. The corresponding Wilson coefficients cannot be generic order one numbers, since the operators are not invariant under the flavor symmetry  $\mathcal{G}_{\text{flavor}}$ . The leading term for the first operator reads

$$(c_1^{\text{VLL}})_{\text{MFV}} \mathcal{O}_1^{\text{VLL}} = Z y_t^4 (V_{ts}^* V_{td})^2 (\bar{s} \gamma_\mu P_L d)^2, \quad (32.43)$$

where  $y_t$  is the SM top Yukawa ( $= m_t/v$ ) and  $Z$  is a (flavor independent) coefficient, generically of  $\mathcal{O}(1)$ . Thanks to the suppression by the small Kobayashi-Maskawa elements  $V_{ts}$  and  $V_{td}$ , the bound on the scale  $\Lambda$  of this operator is relatively weak

$$\Lambda \gtrsim 5 \text{ TeV}, \quad (32.44)$$

to be compared to the bound of  $1.6 \times 10^4$  TeV, as shown in the table above.

The other operators have a much smaller Wilson coefficient as they are suppressed by either the strange Yukawa square ( $\mathcal{O}_1^{\text{SLL}}, \mathcal{O}_2^{\text{SLL}}$ ) or the product of down and strange Yukawas ( $\mathcal{O}_1^{\text{LR}}, \mathcal{O}_2^{\text{LR}}$ ), resulting also in weak bounds on the new physics scale  $\Lambda$ .

This structure can be generalized to any higher dimensional operator mediating a flavor transition. Thus, generically in MFV models, flavor changing operators automatically have their SM-like suppressions, proportional to the same Kobayashi-Maskawa elements and quark masses as in the SM and this can naturally address the new physics flavor puzzle, as the new physics scale of MFV models can be  $\mathcal{O}(1 \text{ TeV})$  without violating flavor physics bounds.

To conclude, the MFV ansatz is remarkably successful in satisfying the constraints from low energy flavor observables. However, it does not address the question: *Why do quark and lepton masses, as well as quark mixing, have such a hierarchical pattern* (SM flavor puzzle), since it simply states that the new physics flavor violation has to have the same structure of the SM flavor violation.

### 32.3.3 EFTs for Rare $B$ Decays

Rare  $B_d$  and  $B_s$  decays based on the  $b \rightarrow s$  flavor changing NC transition are very sensitive to beyond the SM physics, as they are much suppressed in the SM. Recent measurements at the LHC, complementing earlier  $B$ -factory results, have hugely increased the available experimental information on these decays. We will focus on the golden channels: the  $B_s$  and  $B_d$  decays to two muons as they are among the rarest  $B$  decays.

In the SM, these decays are dominated by the  $Z$  penguin and box diagrams involving top quark exchanges. The resulting effective Hamiltonian depends, therefore, on the loop function  $Y(x_t)$ , with  $x_t \equiv m_t^2/m_W^2$  and reads

$$\mathcal{H}_{\text{eff}} = -\frac{G_F}{\sqrt{2}} \frac{\alpha}{\pi \sin^2 \theta} V_{tb}^* V_{ts} Y(x_t) (\bar{b} \gamma_\mu P_L s) (\bar{\mu} \gamma_\mu \gamma_5 \mu) + \text{h.c.} , \quad (32.45)$$

with  $s$  replaced by  $d$  in the case of  $B_d \rightarrow \mu^+ \mu^-$ . Evaluating the two matrix elements of the quark and muon currents leads to the branching ratio,

$$\begin{aligned} \text{BR}(B_s \rightarrow \mu^+ \mu^-) &= \frac{G_F^2}{\pi} \left( \frac{\alpha}{4\pi \sin^2 \theta} \right)^2 |V_{tb}^* V_{ts}|^2 Y^2(x_t) \times \\ &\times m_\mu^2 m_{B_s} \sqrt{1 - \frac{4m_\mu^2}{m_{B_s}^2}} F_{B_s} \tau_{B_s} , \end{aligned} \quad (32.46)$$

and analogously for the  $B_d$  decay. In this equation,  $m_{B_s}$  is the mass of the  $B_s$  meson,  $\tau_{B_s}$  its life time (1.6 ps), and  $F_{B_s}$  the corresponding decay constant.

The main theoretical uncertainties in this branching ratio result from the uncertainties in the decay constant ( $\sim 4\%$  for  $B_d$  and  $\sim 3\%$  for  $B_s$ , using the latest lattice computations) and in the Kobayashi-Maskawa elements  $V_{td}$  and  $V_{ts}$  (both at the level of several %).

Inserting numbers and including the  $\mathcal{O}(\alpha)$  and  $\mathcal{O}(\alpha_s^2)$  corrections, the SM predictions read

$$\begin{aligned} \text{BR}(B_s \rightarrow \mu^+ \mu^-)_{\text{SM}} &= (3.65 \pm 0.23) \times 10^{-9} , \\ \text{BR}(B_d \rightarrow \mu^+ \mu^-)_{\text{SM}} &= (1.06 \pm 0.09) \times 10^{-10} . \end{aligned} \quad (32.47)$$

As shown by (32.46), the tiny branching ratios of these decays in the SM are due to the loop, Kobayashi-Maskawa and helicity (by the small muon mass) suppressions. Extensions of the SM do not necessarily contain any of these suppression mechanisms, in particular the helicity suppression.

Experimentally, searches for  $B_{s,d} \rightarrow \mu^+ \mu^-$  have been performed by 11 experiments, spanning more than three decades. The first hint for a non-zero  $B_s$  decay was reported in 2011 by the CDF collaboration. This was followed by several measurements by ATLAS, CMS and LHCb. Currently we have more than  $6\sigma$  evidence for  $B_s \rightarrow \mu^+ \mu^-$  showing a good agreement with the SM prediction.

Several additional operators can contribute to the  $B_{s,d}$  decays in beyond SM theories:  $\mathcal{O}'_{10}$ , obtained from the SM operator in (32.45) with  $P_L \rightarrow P_R$  and

$$\mathcal{O}_S = (\bar{b}P_L s)(\bar{\mu}\mu) , \quad \mathcal{O}_P = (\bar{b}P_L s)(\bar{\mu}\gamma_5\mu) , \quad (32.48)$$

and the corresponding prime operators obtained by  $P_L \rightarrow P_R$ . Using these additional operators, one can compute the ratio

$$\begin{aligned} \frac{\Gamma(B_s \rightarrow \mu^+ \mu^-)}{\Gamma(B_s \rightarrow \mu^+ \mu^-)_{\text{SM}}} &\simeq (|S_s|^2 + |P_s|^2) \left( \frac{1}{1 + y_s} \right) \times \\ &\times \left( 1 + y_s \frac{\text{Re}(P_s^2) - \text{Re}(S_s^2)}{|S_s|^2 + |P_s|^2} \right) , \end{aligned} \quad (32.49)$$

where  $y_s = (8.8 \pm 1.4)\%$  ( $y_d \sim 0$  for the  $B_d$  system) have to be taken into account when comparing experimental and theoretical results, and

$$S_s \equiv \frac{m_{B_s}}{2m_\mu} \frac{(C_s^S - C_s'^S)}{C_{10s,d}^{SM}} \sqrt{1 - \frac{4m_\mu^2}{m_{B_s}^2}} , \quad (32.50)$$

$$P_s \equiv \frac{m_{B_s}}{2m_\mu} \frac{(C_s^P - C_s'^P)}{C_{10s,d}^{SM}} + \frac{(C_s^{10} - C_{10s}')}{C_{10s}^{SM}} , \quad (32.51)$$

with the several Wilson coefficients defined using the normalization

$$\mathcal{H}_{\text{eff}} = -4 \frac{G_F}{\sqrt{2}} V_{tb} V_{ts}^* \frac{e^2}{16\pi^2} \sum_i (C_i \mathcal{O}_i + C_i' \mathcal{O}_i') + \text{h.c.} . \quad (32.52)$$

Similar expressions hold for the  $B_d$  system. It is evident that the helicity suppression of the branching ratio can be eliminated thanks to the scalar and pseudoscalar operators and, therefore, large enhancements can be obtained. Comparing with the latest measurement of  $B_s \rightarrow \mu^+ \mu^-$ , one can find the bounds on the Wilson coefficients of the scalar and pseudoscalar operators.

As it is well known, the measurement of the ratio between  $\text{BR}(B_s \rightarrow \mu^+ \mu^-)$  and  $\text{BR}(B_d \rightarrow \mu^+ \mu^-)$  gives a very clean probe of new sources of flavor violation beyond the Kobayashi-Maskawa matrix. Indeed, in all MFV models,

the ratio is determined by

$$\frac{\Gamma(B_d \rightarrow \mu^+ \mu^-)}{\Gamma(B_s \rightarrow \mu^+ \mu^-)} = \frac{\tau_{B_d} m_{B_d} F_{B_d}^2 |V_{td}|^2}{\tau_{B_s} m_{B_s} F_{B_s}^2 |V_{ts}|^2} \sim 0.03, \quad (32.53)$$

and has a relatively small theoretical uncertainty at the level of  $\sim 5\%$ . In the coming years LHCb, ATLAS and CMS collaborations will be able to produce a more accurate test of this relation and, therefore, of the MFV ansatz.

### 32.3.4 Top and Higgs Flavor Violating Signatures

So far we have discussed low energy flavor observables that have been/will be measured by  $B$ -factories and by the LHCb. High energy flavor measurements by the ATLAS and CMS collaborations provide a complementary tool to test the underlying flavor structure of Nature. Particularly, in the last few years, a tremendous progress has been achieved in the measurement of Higgs and top flavor violating couplings. This is the topic of this section.

The top quark is the only quark whose Yukawa coupling to the Higgs boson is order of unity and the only one with a mass larger than the mass of the weak gauge bosons. Thanks to its heavy mass, the top mainly decays to a  $W$  boson and a bottom quark, with an extremely small life time of approximately  $5 \times 10^{-25}$  s. This is shorter than the hadronization time, making it impossible for the top quark to form bound states. For these reasons the top quark plays a special role in the SM and in many extensions thereof. An accurate knowledge of its properties can bring key information on fundamental interactions at the EW scale and beyond. So far, the flavor conserving properties of the top are known with a very good accuracy. Less is known about the flavor changing top couplings.

The flavor changing decays of the top quark are suppressed by the GIM mechanism, similarly to what happens to the other quarks. The decay of a top quark to a  $Z$  boson or a photon and an up or charm quark occurs only through higher-order diagrams. These processes should be compared to the tree-level decay to a  $W$  boson and a bottom quark, resulting in tiny top flavor changing branching ratios in the framework of the SM.

In the second column of the table below, we present the SM predictions for the flavor changing branching ratios of the top.

Decay mode	SM prediction	LHC bound	Comments
BR ( $t \rightarrow ch$ )	$3 \times 10^{-15}$	$4.6 \times 10^{-3}$	$h \rightarrow \text{lept.}, h \rightarrow b\bar{b}, \gamma\gamma$
BR ( $t \rightarrow uh$ )	$2 \times 10^{-17}$	$4.2 \times 10^{-3}$	$h \rightarrow b\bar{b}, h \rightarrow \gamma\gamma$
BR ( $t \rightarrow cg$ )	$5 \times 10^{-12}$	$2 \times 10^{-4}$	Single top production
BR ( $t \rightarrow ug$ )	$4 \times 10^{-14}$	$4 \times 10^{-5}$	Single top production
BR ( $t \rightarrow uZ$ )	$8 \times 10^{-17}$	$1.7 \times 10^{-4}$	$Z \rightarrow \ell\ell, tZ$ production
BR ( $t \rightarrow cZ$ )	$10^{-14}$	$2 \times 10^{-4}$	$Z \rightarrow \ell\ell, tZ$ production
BR ( $t \rightarrow u\gamma$ )	$4 \times 10^{-16}$	$1.3 \times 10^{-4}$	Single top production
BR ( $t \rightarrow c\gamma$ )	$5 \times 10^{-14}$	$1.7 \times 10^{-3}$	Single top production

All branching ratios are below the  $10^{-13}$  level! A discovery of a flavor violating top decay in the foreseeable future would, therefore, unequivocally, imply the existence of New Physics.

Several searches for top flavor changing couplings have been performed at the LHC. So far, there is no evidence for non-zero couplings. In the third column of the Table we show the state of the art of the most stringent constraints on the several branching ratios. All searches have been performed using the full 8 TeV luminosity. Some searches look directly for top flavor changing decays; some other for single top production, eventually in association with a  $Z$  or a photon. Projections of these constraints for the LHC show that we could reach the sensitivity to flavor changing branching ratios at the level of

$$\text{BR}(t \rightarrow gc) \lesssim 4 \times 10^{-6}, \quad \text{BR}(t \rightarrow hq) \lesssim 2 \times 10^{-4}. \quad (32.54)$$

These values are still quite larger than the corresponding SM predictions, but will be crucial for testing the prediction of 2HDMs models and of Randall-Sundrum-like models with a generic flavor structure, that can predict branching ratios as large as

$$\begin{aligned} \text{BR}(t \rightarrow gc)_{2\text{HDM}} \sim 10^{-5}, \quad \text{BR}(t \rightarrow hq)_{2\text{HDM}} \sim 2 \times 10^{-3}, \\ \text{BR}(t \rightarrow hq)_{\text{RS}} \sim 10^{-4}, \end{aligned} \quad (32.55)$$

in agreement with the present low energy flavor constraints.

As we have discussed above, the Higgs is intrinsically connected to the flavor puzzle, as without Yukawa interactions the SM flavor symmetry,  $\mathcal{G}_{\text{flavor}}$ , would be un-broken. For this reason, it is of paramount importance to test

the couplings of the Higgs with quarks and leptons at the LHC. By now, we know that the masses of the third generation quarks and leptons are largely due to the 125 GeV Higgs, as indicated by the measured values of Higgs couplings to the third generation fermions. Little is known about the origin of the masses of the first and second generation fermions and about flavor changing Higgs couplings.

In the SM, in spite of the very small Higgs width, flavor violating Higgs decays have a negligible branching ratio. Generically, flavor violating Yukawa couplings are well constrained by the low energy flavor changing NC measurements. A notable exception are the flavor violating couplings involving a tau lepton. Models with extra sources of EW SSB, can predict a sizable (% level) Higgs flavor violating decays to a tau and a lepton, while being in agreement with low energy flavor observables, as  $\tau \rightarrow \mu\gamma$ .

A few searches for Higgs flavor violating decays  $h \rightarrow \tau\mu, \tau e$  have been performed by the LHC, so far not showing a convincing evidence for non-zero branching ratios (see, however, the initial small anomaly. It will be very interesting to monitor these searches in the coming years of the LHC, as they could give a complementary probe of models with sizable flavor changing Higgs couplings to leptons.

**Exercise 32.1:** Prove that the neutral interactions of the photon, the  $Z$  boson, the gluons and the Higgs boson are flavor diagonal in the quark mass eigenbasis.

**Exercise 32.2:** Using the unitarity relations show that the Kobayashi-Maskawa matrix is fully described by 4 free parameters.

**Exercise 32.3:** Write the leading term of the Wilson coefficient of each operator in (32.34), according to the flavor violating *ansatz* and demonstrate that they are much smaller than ( $c_1^{VLL}$ ).

**Exercise 32.4:** Write down the transformations of spin and linear momentum vectors of an elementary particle under the parity and time reversal operators. Suggest a way to look for time reversal violation in the decay  $\Lambda \rightarrow N + \pi$ . Are any experimental details or assumptions crucial to this suggestion?

**Exercise 32.4:** The state of two  $\pi^0$ -mesons is the  $CP$  odd or even?

## Chapter 33

# Neutrino Mass Problem

Discovery of the neutrino oscillations in the atmospheric Super-Kamiokande, solar SNO and reactor KamLAND experiments was a first evidence in favor of a beyond the SM physics. Neutrino oscillations were further studied in the long baseline accelerator K2K, MINOS and T2K experiments. With the measurement of the small parameter  $\sin^2 \theta_{13}$  (in the accelerator T2K, reactor Daya Bay, RENO and Double Chooze experiments) investigations of neutrino oscillations enters into a new era, era of high precision measurements.

The 2015 Nobel Prize to Kajita and McDonald for the discovery of neutrino oscillations, which shows that neutrinos have mass, is a very important event for the neutrino community that will attract new people and give a great boost to the field.

### 33.1 Brief Historical Survey

Idea of neutrino oscillations was first proposed by Pontecorvo in 1957-58 soon after the theory of the two-component neutrino was proposed and confirmed by the Goldhaber et al. Pontecorvo looked in the lepton world for a phenomena analogous to  $K^0 \rightleftharpoons \bar{K}^0$  oscillations. At that time only one type of neutrino was known. Pontecorvo assumed that in addition to the usual weak interaction exist a much weaker interaction which does not conserve the lepton number. Assuming maximum mixing (by the analogy with  $K^0 - \bar{K}^0$ ) he concluded that "...neutrino and antineutrino are particle mixtures, i.e. sym-

metrical and antisymmetrical combinations of two truly neutral Majorana particles  $\nu_1$  and  $\nu_2$ ...":

$$|\bar{\nu}_R\rangle = \frac{1}{\sqrt{2}}(|\nu_1\rangle + |\nu_2\rangle), \quad |\nu_R\rangle = \frac{1}{\sqrt{2}}(|\nu_1\rangle - |\nu_2\rangle). \quad (33.1)$$

Here  $|\bar{\nu}_R\rangle$  is the state of the right antineutrino, while  $|\nu_R\rangle$  is the state of the right neutrino, a particle which does not take part in the weak interaction (later Pontecorvo proposed the name sterile for such neutrinos), and  $|\nu_{1,2}\rangle$  are states of Majorana neutrinos with small masses  $m_{1,2}$ . As a result of the mixing (33.1), oscillations  $\bar{\nu}_R \leftrightarrow \nu_R$  (sterile) become possible.

Pontecorvo discussed a possibility of experimental confirmation of the neutrino oscillation hypothesis. In 1958 the only known sources of neutrinos were reactors and the sun. Pontecorvo finished his paper with the remark: "... effects of transformation of neutrino into antineutrino and vice versa may be unobservable in the laboratory because of large values of  $R$  (oscillation length), but will certainly occur, at least, on an astronomic scale".

In 1962 the idea of neutrino masses and mixing was discussed by Maki, Nakagawa and Sakata. Their proposal was based on the Nagoya model in which nucleons were considered as bound states of a vector boson and neutrino with definite mass. They assumed that the fields of the weak neutrinos  $\nu_e$  and  $\nu_\mu$  are connected with the fields of neutrinos with definite masses  $\nu_1$  and  $\nu_2$  (they called them true neutrinos) by the orthogonal transformation

$$\nu_e = \cos\theta\nu_1 + \sin\theta\nu_2, \quad \nu_\mu = -\sin\theta\nu_1 + \cos\theta\nu_2. \quad (33.2)$$

The phenomenon of neutrino oscillations was not considered, however, a possibility of 'virtual transmutation' of  $\nu_\mu$  into  $\nu_e$  was discussed. They estimated a time of this transition and discussed how a possible  $\nu_\mu \rightarrow \nu_e$  transition would influence the interpretation of the results of the Brookhaven experiment which was going on the time when this paper was written.

In 1967 Pontecorvo published the second paper on neutrino oscillations. In this paper he discussed flavor neutrino oscillations  $\nu_\mu \leftrightarrow \nu_e$  and also oscillations between flavor and sterile neutrinos ( $\nu_{eL} \leftrightarrow \bar{\nu}_{eL}$  etc.). In this paper solar neutrino oscillations were considered. Before the first results of the Davis solar neutrino experiment appeared, Pontecorvo pointed out that because of neutrino oscillations the flux of the solar  $\nu_e$ 's could be two times smaller than the expected flux. Thus, he anticipated *the solar neutrino problem*.



In the Gribov and Pontecorvo paper it was suggested that only active left-handed neutrinos  $\nu_e$  and  $\nu_\mu$  and right-handed antineutrinos  $\bar{\nu}_e$  and  $\bar{\nu}_\mu$  exist in nature (no sterile neutrinos). It was assumed that exist a (mili-weak) interaction which does not conserve lepton numbers. After the diagonalization of such an interaction the authors came to the mixing relation

$$\nu_{eL} = \cos \xi \phi_{1L} + \sin \xi \phi_{2L} , \quad \nu_{\mu L} = -\sin \xi \phi_{1L} + \cos \xi \phi_{2L} , \quad (33.3)$$

where  $\xi$  is the mixing angle and  $\phi_1$  and  $\phi_2$  are fields of the Majorana neutrinos with masses  $m_1$  and  $m_1$ . They calculated the probability of  $\nu_e$  to survive in vacuum. The case of the maximum mixing ( $\xi = \pi/4$ ), analogous to the  $K^0 - \bar{K}^0$  case, was considered as the most attractive one. Under this assumption the oscillations of solar neutrinos were discussed.

In the 1970-80-ies an idea of neutrino masses and oscillations was further developed in Dubna. In addition to the Gribov-Pontecorvo scheme of the neutrino mixing, based on the Majorana mass term, neutrino mixing based on the Dirac mass term and the most general Dirac and Majorana mass term were considered. Possible reactor, accelerator, solar and atmospheric experiments on the search for neutrino oscillations were discussed.

Currently majority of physicist believe that:

- There are no principles which require that neutrinos are massless particles. It is plausible that neutrinos have small nonzero masses;
- Neutrino oscillations is an interference phenomenon. Search for neutrino oscillations is the most sensitive method to search for extremely small mass-squared differences;
- Experiments with neutrinos from different sources are sensitive to different neutrino mass-squared differences. Experiments on the search for neutrino oscillations must be performed with neutrinos from all existing sources.

## 33.2 Neutrino Mixing

Neutrino oscillations are based on the mixing of neutrino fields

$$\nu_{iL}(x) = \sum_i U_{li} \nu_{iL}(x) , \quad (33.4)$$

where  $U$  is a unitary mixing matrix and  $\nu_i(x)$  is the field of neutrinos (Dirac or Majorana) with mass  $m_i$ .

The flavor neutrino fields  $\nu_{lL}(x)$  ( $l = e, \mu, \tau$ ) enter into the SM CC and NC interactions

$$\mathcal{L}_I^{CC} = -\frac{g}{\sqrt{2}} j_\alpha^{CC} W^\alpha + \text{h.c.} , \quad \mathcal{L}_I^{NC} = -\frac{g}{2 \cos \theta_W} j_\alpha^{NC} Z^\alpha . \quad (33.5)$$

Here

$$j_\alpha^{CC} = \sum_{l=e,\mu,\tau} \bar{\nu}_{lL} \gamma_\alpha l_L , \quad j_\alpha^{NC} = \sum_{l=e,\mu,\tau} \bar{\nu}_{lL} \gamma_\alpha \nu_{lL} \quad (33.6)$$

are charged leptonic and neutral neutrino currents.

The neutrino mixing takes place if in the total Lagrangian there is a mass term, nondiagonal over flavor neutrino fields. In the case of the charged particles (leptons and quarks) only Dirac mass terms are possible. Because the electric charges of neutrinos are equal to zero three different neutrino mass terms considered.

- **Dirac Mass Term,**

$$\mathcal{L}^D = - \sum_{l=e,\mu,\tau} \bar{\nu}_{lL} M_{l,l}^D \nu_{lR} + \text{h.c.} , \quad (33.7)$$

where  $M^D$  is a complex, nondiagonal,  $3 \times 3$  matrix. After the diagonalization of the matrix  $M^D$  we have

$$\nu_{lL}(x) = \sum_{i=1}^3 U_{li} \nu_{iL}(x) . \quad (33.8)$$

Here  $U$  is the unitary mixing matrix and  $\nu_i(x)$  is the Dirac field with the mass  $m_i$ . The Lagrangian  $\mathcal{L}^D$  conserves the total lepton number

$$L = L_e + L_\mu + L_\tau . \quad (33.9)$$

Neutrino  $\nu_i$  and antineutrino  $\bar{\nu}_i$  differ by the lepton number:

$$L(\nu_i) = 1 , \quad L(\bar{\nu}_i) = -1 . \quad (33.10)$$

- **Majorana Mass Term,**

$$\mathcal{L}^M = -\frac{1}{2} \sum_{l=e,\mu,\tau} \bar{\nu}_{lL} M_{ll}^M (\nu_{lL})^c + \text{h.c.} , \quad (33.11)$$

where  $M^M$  is a complex, nondiagonal, symmetrical  $3 \times 3$  matrix and

$$(\nu_{iL})^c = C \bar{\nu}_{iL}^T \quad (33.12)$$

is the conjugated field. The mass term (33.11) violates not only flavor lepton numbers, but also the total lepton number  $L$ . After the diagonalization of the matrix  $M^M$  we have

$$\nu_{iL}(x) = \sum_{i=1}^3 U_{li} \nu_{iL}(x) . \quad (33.13)$$

Here  $U$  is a unitary  $3 \times 3$  mixing matrix and

$$\nu_i(x) = \nu_i^c(x) \quad (33.14)$$

is the Majorana field with the mass  $m_i$  ( $\nu_i \equiv \bar{\nu}_i$ ).

- The most general **Dirac and Majorana mass term**,

$$\mathcal{L}^{D+M} = \mathcal{L}^M + \mathcal{L}^D - \sum_{s', s=s_1, \dots, s_{n_s}} \frac{1}{2} \overline{(\nu_{s'R})^c} M_{s's}^R \nu_{sR} + \text{h.c.} \quad (33.15)$$

( $M^R$  is a complex symmetrical matrix), violates lepton numbers and require left-handed and right-handed neutrino fields. After the diagonalization of the mass term  $\mathcal{L}^{D+M}$  we find

$$\nu_{iL}(x) = \sum_{i=1}^{3+n_s} U_{li} \nu_{iL}(x) , \quad (\nu_{sR}(x))^c = \sum_{i=1}^{3+n_s} U_{si} \nu_{iL}(x) . \quad (33.16)$$

Here  $U$  is a unitary  $(3 + n_s) \times (3 + n_s)$  matrix and

$$\nu_i(x) = \nu_i^c(x) \quad (33.17)$$

is the field of a Majorana lepton with definite mass.

The mixing (33.16) open different possibilities: the seesaw possibility of the generation of small neutrino masses, a possibility of transitions of flavor neutrinos into sterile states etc.

Let us notice that the Dirac mass term can be generated by the standard Higgs mechanism. The Majorana and the Dirac-Majorana mass terms can be generated only by a beyond the SM mechanisms.

### 33.3 Flavor Neutrino States

There exist different methods of the derivation of the discussed above expression for transition probabilities. We will present here a method based on the notion of the coherent flavor neutrino states

$$|\nu_l\rangle = \sum_i U_{li}^* |\nu_i\rangle . \quad (l = e, \mu, \tau) \quad (33.18)$$

Here  $|\nu_i\rangle$  is the state of neutrino (Dirac or Majorana) with mass  $m_i$ , momentum  $\vec{p}$  and energy

$$E_i = \sqrt{p^2 + m_i^2} \simeq E + \frac{m_i^2}{2E} \quad (E = p) , \quad (33.19)$$

and  $|\nu_l\rangle$  is the state the flavor neutrino  $\nu_l$  which is produced together with  $l^+$  in a CC weak decay ( $\pi^+ \rightarrow \mu^+ + \nu_\mu$ , etc.) or produces  $l^-$  in a CC neutrino reaction ( $\nu_\mu + N \rightarrow \mu^- + X$ , etc.).

The relation (33.18) is valid if neutrino mass-squared differences are so small that in weak decays production of neutrinos with different masses cannot be resolved. It follows from the Heisenberg uncertainty relation that this condition is satisfied in neutrino oscillation experiments with neutrino energies many orders of magnitude larger than neutrino masses.

The possibility to resolve small neutrino mass-squared differences is based on the time-energy uncertainty relation

$$\Delta E \Delta t \gtrsim 1 . \quad (33.20)$$

Here  $\Delta t$  is a time interval during which the state with the energy uncertainty  $\Delta E$  is significantly changed. In the case of neutrino beams, from (33.20) we find

$$|\Delta m_{ki}^2| \frac{L}{2E} \gtrsim 1 , \quad (33.21)$$

where  $L \simeq \Delta t$  is the distance between a neutrino source and neutrino detector. For 'atmospheric' and 'solar' mass-squared differences,

$$\Delta m_A^2 \simeq 2.4 \cdot 10^{-3} \text{ eV}^2 , \quad \Delta m_S^2 \simeq 7.5 \cdot 10^{-5} \text{ eV}^2 , \quad (33.22)$$

the condition (33.21) is satisfied in the atmospheric Super-Kamiokande, long baseline accelerator K2K, MINOS, T2K, reactor KamLAND, Daya Bay, RENO, Double Chooze and other neutrino oscillation experiments.

Let us finish this section with a remark about the states of sterile neutrinos which (by definition) do not interact with leptons and quarks via the SM interaction. If in addition to the flavor neutrinos  $\nu_l$  sterile neutrinos  $\nu_s$  exist, their states are determined as follows

$$|\nu_s\rangle = \sum_{i=1}^{3+n_s} U_{si}^* |\nu_i\rangle, \quad (s = s_1, s_1, \dots) \quad (33.23)$$

where  $U$  is a unitary  $(3 + n_s) \times (3 + n_s)$  matrix. Then the states of active and sterile neutrinos (33.18) and (33.23) satisfy the condition

$$\langle \alpha' | \alpha \rangle = \delta_{\alpha' \alpha}. \quad (\alpha', \alpha = e, \mu, \tau, s_1, s_1, \dots, s_{n_s}) \quad (33.24)$$

Neutrino oscillations is a direct consequence of the fact that flavor (and sterile) neutrinos are described by coherent states (33.18) and (33.23).

### 33.4 Neutrino Oscillations in Vacuum

Let us assume that at the initial time  $t = 0$  a flavor neutrino  $\nu_\alpha$  was produced. In the general case of flavor and sterile neutrinos at the time  $t$  we have

$$\begin{aligned} |\nu_\alpha\rangle_t &= e^{-iH_0 t} |\nu_\alpha\rangle = \sum_{i=1}^{3+n_s} |\nu_i\rangle e^{-iE_i t} U_{\alpha i}^* = \\ &= \sum_{\alpha'} |\alpha'\rangle \left( \sum_{i=1}^{3+n_s} U_{\alpha' i} e^{-iE_i t} U_{\alpha i}^* \right). \end{aligned} \quad (33.25)$$

Thus, for the  $\nu_\alpha \rightarrow \nu_{\alpha'}$  transition probability we find

$$P(\nu_\alpha \rightarrow \nu_{\alpha'}) = \left| \sum_{i=1}^{3+n_s} U_{\alpha' i} e^{-iE_i t} U_{\alpha i}^* \right|^2. \quad (33.26)$$

We will present here convenient expression for  $\nu_\alpha \rightarrow \nu_{\alpha'}$  transition probability. From (33.26) we have

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_{\alpha'}) &= \left| \sum_{i=1}^{3+n_s} U_{\alpha' i} e^{-2i\Delta_{pi}} U_{\alpha i}^* \right|^2 = \\ &= \left| \delta_{\alpha' \alpha} - 2i \sum_i U_{\alpha' i} e^{-i\Delta_{pi}} \sin \Delta_{pi} U_{\alpha i}^* \right|^2, \end{aligned} \quad (33.27)$$

where  $p$  is arbitrary fixed index and

$$\Delta_{pi} = \frac{\Delta m_{pi}^2 L}{4E}, \quad \Delta m_{pi}^2 = m_i^2 - m_p^2. \quad (33.28)$$

Let us notice that in (33.27)  $i \neq p$ , also we extracted the common phase  $e^{-im_p^2 L/2E}$  and used the unitarity condition  $\sum_i U_{\alpha'i} U_{\alpha i}^* = \delta_{\alpha'\alpha}$ . From (33.27) we find

$$\begin{aligned} P(\nu\alpha \rightarrow \nu\alpha') &= \delta_{\alpha'\alpha} - 4 \sum_i |U_{\alpha i}|^2 (\delta_{\alpha'\alpha} - |U_{\alpha' i}|^2) \sin^2 \Delta_{pi} + \\ &+ 8 \sum_{i>k} [\text{Re}(U_{\alpha' i} U_{\alpha i}^* U_{\alpha' k}^* U_{\alpha k}) \cos(\Delta_{pi} - \Delta_{pk}) \pm \\ &\pm \text{Im}(U_{\alpha' i} U_{\alpha i}^* U_{\alpha' k}^* U_{\alpha k}) \sin(\Delta_{pi} - \Delta_{pk})] \sin \Delta_{pi} \sin \Delta_{pk}. \end{aligned} \quad (33.29)$$

Here  $+$  ( $-$ ) sign refer to  $\nu\alpha \rightarrow \nu\alpha'$  ( $\bar{\nu}\alpha \rightarrow \bar{\nu}\alpha'$ ) transition. There are some advantages of (33.29) since it contains only independent mass-squared differences. The unitarity condition is fully implemented in (33.29) and as a result only independent terms enter into this expression.

Let us consider now the most important case of the three-neutrino mixing. Usually neutrino masses are labeled in such a way that  $m_2 > m_1$  and solar ('small') mass-squared difference is determined as follows

$$m_2^2 - m_1^2 = \Delta m_{12}^2 \equiv \Delta m_S^2. \quad (33.30)$$

For the neutrino mass spectrum there are two possibilities:

1. Normal spectrum (NS):  $\Delta m_S^2$  is the difference between square of masses of the lightest neutrinos. In this case  $m_3 > m_2 > m_1$ ;
2. Inverted spectrum (IS):  $\Delta m_S^2$  is the difference between square of masses of the heaviest neutrinos. In this case  $m_2 > m_1 > m_3$ .

We will determine the atmospheric ('large') neutrino mass squared difference in the following way

$$\text{NS} : \Delta m_A^2 = \Delta m_{23}^2, \quad \text{IS} : \Delta m_A^2 = |\Delta m_{13}^2|. \quad (33.31)$$

Let us notice that there exist different definition of this quantity in the literature:

- In some cases atmospheric mass-squared difference is determined as:

$$(\Delta m_A^2)' = \frac{1}{2} |\Delta m_{13}^2 + \Delta m_{23}^2| = \Delta m_A^2 + \frac{1}{2} \Delta m_S^2 . \quad (33.32)$$

- Sometimes atmospheric mass-squared difference is determined in the following way:

$$(\Delta m_A^2)'' = \Delta m_{13}^2 \text{ (NS)} = |\Delta m_{23}^2| \text{ (IS)} = \Delta m_A^2 + \Delta m_S^2 . \quad (33.33)$$

- The parameter  $\Delta m_{ee}^2$  is expressed as follows:

$$\Delta m_{ee}^2 = \cos^2 \theta_{12} \Delta m_{13}^2 + \sin^2 \theta_{12} \Delta m_{23}^2 , \quad (33.34)$$

and is connected with  $\Delta m_A^2$  and  $\Delta m_S^2$  by the relations

$$\begin{aligned} \Delta m_{ee}^2 &= \Delta m_A^2 + \cos^2 \theta_{12} \Delta m_S^2 \text{ (NS)} , \\ |\Delta m_{ee}^2| &= \Delta m_A^2 + \sin^2 \theta_{12} \Delta m_S^2 \text{ (IS)} . \end{aligned} \quad (33.35)$$

As it is seen from (33.31), (33.32), (33.33) and (33.35) different definitions of 'large' mass-squared difference differ only by a few %. However, neutrino oscillation experiments enter now into precision era when neutrino oscillation parameters will be measured with 1 % accuracy. It is believed that the consensus in definition of 'large' neutrino mass-squared difference must be found.

For the probability of the transition  $\nu l \rightarrow \nu l'$  ( $l, l' = e, \mu, \tau$ ) in the case of normal and inverted mass spectra from (33.29) we find, correspondingly, the following expressions

$$\begin{aligned} P^{\text{NS}}(\nu l \rightarrow \nu l') &= \delta_{\nu l} - 4|U_{l3}|^2(\delta_{\nu l} - |U_{\nu 3}|^2) \sin^2 \Delta_A - \\ &- 4|U_{l1}|^2(\delta_{\nu l} - |U_{\nu 1}|^2) \sin^2 \Delta_S - \\ &- 8[\text{Re}(U_{\nu 3} U_{l3}^* U_{\nu 1}^* U_{l1}) \cos(\Delta_A + \Delta_S) \pm \\ &\pm \text{Im}(U_{\nu 3} U_{l3}^* U_{\nu 1}^* U_{l1}) \sin(\Delta_A + \Delta_S)] \sin \Delta_A \sin \Delta_S , \end{aligned} \quad (33.36)$$

and

$$\begin{aligned} P^{\text{IS}}(\nu l \rightarrow \nu l') &= \delta_{\nu l} - 4|U_{l3}|^2(\delta_{\nu l} - |U_{\nu 3}|^2) \sin^2 \Delta_A - \\ &- 4|U_{l2}|^2(\delta_{\nu l} - |U_{\nu 2}|^2) \sin^2 \Delta_S - \\ &- 8[\text{Re}(U_{\nu 3} U_{l3}^* U_{\nu 2}^* U_{l2}) \cos(\Delta_A + \Delta_S) \mp \\ &\mp \text{Im}(U_{\nu 3} U_{l3}^* U_{\nu 2}^* U_{l2}) \sin(\Delta_A + \Delta_S)] \sin \Delta_A \sin \Delta_S . \end{aligned} \quad (33.37)$$

The transition probabilities (33.37) and (33.38) are the sum of atmospheric, solar and interference terms. Notice that expression (33.38) can be obtained from (33.37) by the change  $U_{l1} \rightarrow U_{l2}$  and  $(\pm) \rightarrow (\mp)$  in the last term.

The values of the oscillation parameters obtained from global analysis of existing data are presented in the Table:

Parameter	Normal Spectrum	Inverted Spectrum
$\sin^2 \theta_{12}$	$0.304^{+0.013}_{-0.012}$	$0.304^{+0.013}_{-0.012}$
$\sin^2 \theta_{23}$	$0.452^{+0.052}_{-0.028}$	$0.579^{+0.025}_{-0.037}$
$\sin^2 \theta_{13}$	$0.0218^{+0.0010}_{-0.0010}$	$0.0219^{+0.0011}_{-0.0010}$
$\delta$ (in $^\circ$ )	$(306^{+39}_{-70})$	$(254^{+63}_{-62})$
$\Delta m_S^2$	$(7.50^{+0.19}_{-0.17}) \cdot 10^{-5} \text{ eV}^2$	$(7.50^{+0.19}_{-0.17}) \cdot 10^{-5} \text{ eV}^2$
$\Delta m_A^2$	$(2.457^{+0.047}_{-0.047}) \cdot 10^{-3} \text{ eV}^2$	$(2.449^{+0.048}_{-0.047}) \cdot 10^{-3} \text{ eV}^2$

### 33.5 Neutrino and SM

The SM started with the theory of the two-component neutrino. The two-component, massless, Weil neutrino is the simplest possibility for the particle with spin 1/2: only two degrees of freedom. The local  $SU_L(2)$  group with the lepton doublets

$$\psi_{eL}^{lep} = \begin{pmatrix} \nu'_{eL} \\ e'_L \end{pmatrix}, \quad \psi_{\mu L}^{lep} = \begin{pmatrix} \nu'_{\mu L} \\ \mu'_L \end{pmatrix}, \quad \psi_{\tau L}^{lep} = \begin{pmatrix} \nu'_{\tau L} \\ \tau'_L \end{pmatrix} \quad (33.38)$$

and corresponding quark doublets is the simplest possibility which allows to include charged leptons and quarks in addition to neutrinos.

In order to unify weak and electromagnetic interactions we need to enlarge the symmetry group: in electromagnetic currents of charged particles enter left-handed and right-handed fields. The simplest enlargement is the  $SU_L(2) \times U_Y(1)$  group where  $U_Y(1)$  is the group of the weak hypercharge  $Y$  determined by the Gell-Mann-Nishijima relation

$$Q = T_3 + \frac{1}{2}Y. \quad (33.39)$$

Neutrinos have no electromagnetic interaction. Unification of the weak and electromagnetic interactions does not require right-handed neutrino fields.



The SM interaction of leptons, neutrinos and quarks with gauge vector bosons *is the minimal interaction* compatible with the local  $SU_L(2) \times U_Y(1)$  invariance.

The SM mechanism of the mass generation is the Higgs mechanism based on the assumption of the existence of scalar Higgs fields. In order to generate masses of  $W^\pm$  and  $Z^0$  bosons we need to have three (Goldstone) degrees of freedom. Minimal possibility is a doublet of complex Higgs fields (four degrees of freedom). With this assumption one scalar, neutral Higgs boson is predicted. This prediction is in a good agreement with existing LHC data.

Masses of  $W^\pm$  and  $Z^0$  bosons are given in the SM by the relations

$$m_W = \frac{1}{2}g v, \quad m_Z = \frac{1}{2}\sqrt{g^2 + g'^2} v = \frac{g}{2 \cos \theta_W} v, \quad (33.40)$$

where

$$v = (\sqrt{2}G_F)^{-1/2} = 246 \text{ GeV} \quad (33.41)$$

is the parameter which characterizes the scale of the electroweak symmetry breaking. Lepton and quark masses and mixing are due to  $SU_L(2) \times U_Y(1)$  invariant Yukawa interactions which generate Dirac mass terms. For the charged leptons we have

$$\mathcal{L}_Y^{lep} = - \sum_l m_l \bar{l} l, \quad (33.42)$$

where  $m_l = y_l v$  and  $y_l$  is the Yukawa constant. Neutrinos in the minimal SM after spontaneous breaking of the EW symmetry remain two-component, massless, Weyl particles.

## 33.6 The Weinberg Mechanism of Mass Generation

In the framework of the minimal SM neutrino masses and mixing can be generated only by a beyond the SM mechanism. The most general method which allows to describe effects of a beyond the SM physics is the method of the effective Lagrangian. The effective Lagrangian is a  $SU_L(2) \times U_Y(1)$  invariant, dimension five or more local operator built from SM fields. In order to build the effective Lagrangian which generate a neutrino mass term we must use the lepton doublets (33.38) and the Higgs doublet

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}. \quad (33.43)$$

The effective Lagrangian which generate the neutrino mass term has the form

$$\mathcal{L}_I^{\text{eff}} = -\frac{1}{\Lambda} \sum_{l_1, l_2} \left( \bar{\psi}_{l_1 L}^{lep} \tilde{\phi} \right) Y_{l_1 l_2} \left( \tilde{\phi}^T (\psi_{l_2 L}^{lep})^c \right) + \text{h.c.} , \quad (33.44)$$

where the parameter  $\Lambda$  characterizes a scale of a beyond the SM physics ( $\Lambda \gg v$ ) and  $\tilde{\phi} = i\sigma_2 \phi^*$  is the conjugated doublet. Let us stress that the Lagrangian (33.44) does not conserve the total lepton number, it can be generated (in the second order of the perturbation theory) by the seesaw interaction of the Higgs-lepton pair with a heavy Majorana right-handed lepton.

After SSB from (33.44) we come to the Majorana mass term

$$\mathcal{L}^{\text{M}} = -\frac{v^2}{2\Lambda} \sum_{l_1, l_2} \bar{\nu}'_{l_1 L} Y_{l_1 l_2} (\nu'_{l_2 L})^c + \text{h.c.} = -\frac{1}{2} \sum_{i=1}^3 m_i \bar{\nu}_i \nu_i . \quad (33.45)$$

Here  $\nu_i = \nu_i^c$  is the field of the Majorana neutrino with the mass

$$m_i = \frac{v^2}{\Lambda} y_i = \frac{v}{\Lambda} (y_i v) , \quad (33.46)$$

where  $y_i$  is a Yukawa coupling. In (33.46)  $y_i v$  is a 'typical' fermion mass in SM. Thus, neutrino masses, generated by the effective Lagrangian (33.44), are suppressed with respect to 'SM masses' by a factor

$$\frac{v}{\Lambda} = \frac{\text{scale of SM}}{\text{scale of a new physics}} \ll 1 . \quad (33.47)$$

The mechanism we have considered (Weinberg mechanism) is apparently the most economical and natural beyond the SM mechanism of the neutrino mass generation. There are two general consequences of this mechanism.

- Neutrinos with definite masses  $\nu_i$  are Majorana particles;
- The number of neutrinos with definite masses is equal to the number of lepton-quark generations (three). This means that in this scheme there are no transitions of flavor neutrinos into sterile states.

The study of the lepton number violating neutrinoless double  $\beta$ -decay ( $0\nu\beta\beta$ -decay),

$$(A, Z) \rightarrow (A, Z + 2) + e^- + e^- , \quad (33.48)$$

of some even-even nuclei is most sensitive way to investigate the Majorana nature of neutrinos with definite masses. The probability of the process (33.48) is proportional to square of the Majorana neutrino mass,

$$m_{\beta\beta} = \sum_i U_{ei}^2 m_i, \quad (33.49)$$

and is very small. It has the following general form:

$$\frac{1}{T_{1/2}^{0\nu}} = |m_{\beta\beta}|^2 |M^{0\nu}|^2 G^{0\nu}(Q, Z). \quad (33.50)$$

Here  $M^{0\nu}$  is the nuclear matrix element and  $G^{0\nu}(Q, Z)$  is known phase factor.

Several experiments on the search for the  $0\nu\beta\beta$  of different nuclei are going on and are in preparation. Up to now the process was not observed. From the data of recent experiments the following upper bound was obtained

$$|m_{\beta\beta}| < 1.4 \cdot 10^{-1} \text{ eV}. \quad (33.51)$$

In future experiments on the search for  $0\nu\beta\beta$  decay the values  $|m_{\beta\beta}| \simeq$  a few  $\cdot 10^{-2}$  eV are planned to be reached.

Indications in favor of transitions of flavor neutrinos into sterile states were obtained in the LSND and MiniBooNE short baseline accelerator experiments and in the GALLEX and SAGE calibration experiments and in short baseline reactor experiments which were reanalyzed with a new reactor antineutrino flux. Many new short baseline source, reactor and accelerator neutrino experiments on the search for sterile neutrinos with masses  $\sim 1$  eV are in preparation. Hopeful in a few years the sterile neutrino anomaly will be resolved.

**Exercise 33.1:** Assume the neutrino mass is exactly zero. Does the neutrino have a magnetic moment? Along what direction(s) does the neutrino spin point? Along what direction(s) does the antineutrino spin point?

**Exercise 33.2:** What is the velocity of a 3 K neutrino in the universe if the neutrino mass is 0.1 eV?

**Exercise 33.3:** A sensitive way to determine the mass of  $\nu_e$  is to measure:

- (a) The angular distribution in electron-neutrino scattering;
- (b) The electron energy spectrum in beta-decay;
- (c) The neutrino flux from the sun.

**Exercise 33.4:** An experiment in a gold mine in South Dakota has been carried out to detect solar neutrinos using the reaction  $\nu + Cl^{37} \rightarrow Ar^{37} + e^-$ . The detector contains approximately  $4 \times 10^5$  liters of tetrachlorethylene ( $CCl_4$ ). Estimate how many atoms of  $Ar^{37}$  would be produced per day.

**Exercise 33.5:** A neutrino of type  $\nu_e$  is produced at rest at time zero. What is the probability (as a function of time) that the neutrino will be in each of the other states?

**Exercise 33.6:** An experiment to detect neutrino oscillations is being performed. The flight path of the neutrinos is 2,000 meters. Their energy is 100 GeV. The sensitivity is such that the presence of 1% of neutrinos of one type different from that produced at the start of flight path can be measured with confidence. Let us suppose that there is a small perturbing interaction between neutrino types, in the absence of which all three types have the same nonzero rest mass  $m$ . Let the matrix element of this perturbation have the same real value  $E$  between each pair of neutrino types. Take  $m$  to be 20 eV. What is the smallest value of  $E$  that can be detected? How does this depend on  $m$ ?

## Part XII

### Lecture – Supersymmetry



## Chapter 34

# Overview and Current Status

*Supersymmetry* (SUSY) is a principle that proposes a relationship between two basic classes of elementary particles: bosons, which have an integer-valued spin, and fermions, which have a half-integer spin.

SUSY is a possible candidate for a "new physics" and seen as an elegant solution to many current problems in particle physics. If confirmed, it could resolve various areas where current theories are believed to be incomplete. It is believed that a SUSY extension to the SM would resolve major hierarchy problems within gauge theory, by guaranteeing that quadratic divergences of all orders will cancel out in QFT.

In SUSY, each particle from one group would have an associated particle in the other, which is known as its *superpartner*, the spin of which differs by a half-integer. These superpartners would be new and undiscovered particles. For example, there would be a particle called a "selectron" (superpartner electron), a bosonic partner of the electron. In the simplest SUSY theories, with perfectly *unbroken SUSY*, each pair of superpartners would share the same mass and internal quantum numbers besides spin.

If SUSY exists then it consists of a SSB allowing superpartners to differ in mass. *Spontaneously broken SUSY* could solve many mysterious problems in particle physics including the hierarchy problem.

There is no evidence at this time to show whether or not SUSY is correct, or what other extensions to current models might be more accurate. In part this

is because it is only since around 2010 that particle accelerators specifically designed to study physics beyond the SM have become operational. Also it is not yet known where exactly to look for SUSY particles, nor the energies required for their successful search.

The main reasons for SUSY being supported by physicists is that the current theories are known to be incomplete and their limitations are well established, and SUSY would be an attractive solution to some of the major concerns. Direct confirmation would entail production of superpartners in collider experiments, such as the LHC. Indirect methods include the search for a permanent *electric dipole moment* in the known particles, which can arise when the SM particle interacts with the SUSY particles. The current best constraint on the electron electric dipole moment put it to be smaller than  $10^{-28}$  e-cm, equivalent to a sensitivity to new physics at the TeV scale and matching that of the current best particle colliders. An electric dipole moment in any fundamental particle points towards time-reversal violating physics, and therefore also *CP*-symmetry violation via the *CPT* theorem. Such electric dipole moment experiments are also much more scalable than conventional particle accelerators and offer a practical alternative to detecting physics beyond the SM as accelerator experiments become increasingly costly and complicated to maintain.

### 34.1 History of SUSY Invention

As the father of SUSY can be considered Gödel, since he theorized two things that support this fact:

1. He assumed that the universe is not expanding but rotating. The centrifugal force arising from the rotation was what kept everything from collapsing under the force of gravity. What makes this universe weird is the way it mixes up space and time;
2. If the universe is truly supersymmetric, then by completing a sufficiently long round trip in a sufficiently advanced spaceship a resident of Gödel's theoretical Supersymmetric universe could travel to any point in time.

Einstein followed Gödel's theory up by stating: "This separation between



past, present, and future is only an illusion, if a stubborn one".

Gervais and Sakita (in 1971), Golfand and Likhtman (also in 1971), and Volkov and Akulov (1972), independently rediscovered SUSY in the context of QFT. They introduced a radically new type of symmetry of spacetime and fundamental fields, which establishes a relationship between elementary particles of different quantum nature, bosons and fermions, and unifies spacetime and internal symmetries of microscopic phenomena.

Note that SUSY with a consistent Lie-algebraic graded structure (on which the Gervais-Sakita rediscovery was based) first arose in 1971 in the context of an early version of string theory by Ramond, Schwarz and Neveu.

Wess and Zumino (in 1974) identified the characteristic renormalization features of four-dimensional SUSY field theories, which identified them as remarkable QFTs, and they and Salam and their fellow researchers introduced early particle physics applications.

The first realistic SUSY version of the SM was proposed in 1977 by Fayet and is known as the *Minimal Supersymmetric SM* (MSSM).

## 34.2 Motivations

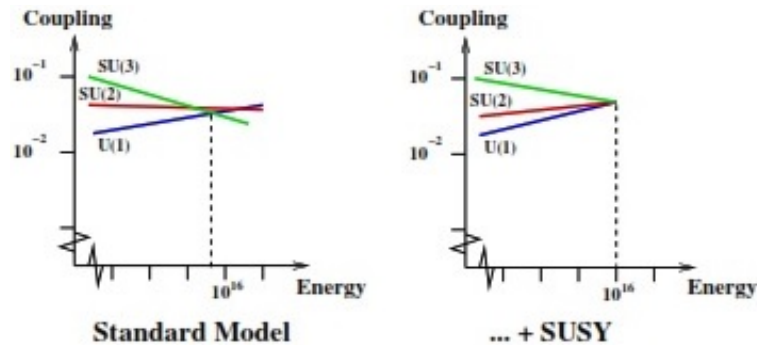
There are numerous phenomenological motivations for SUSY close to the EW scale, as well as technical motivations for SUSY at any scale.

- *The hierarchy problem.* SUSY close to the EW scale ameliorates the hierarchy problem that afflicts the SM.

In the SM, the EW scale receives enormous Planck-scale quantum corrections. The observed hierarchy between the EW scale and the Planck scale must be achieved with extraordinary fine tuning. In a SUSY theory, on the other hand, Planck-scale quantum corrections cancel between partners and superpartners (owing to a minus sign associated with fermionic loops). The hierarchy between the EW scale and the Planck scale is achieved in a natural manner, without miraculous fine-tuning;

- *Gauge coupling unification.* The idea that the gauge symmetry groups unify at high-energy is called Grand Unification Theory (GUT). In

the SM, however, the weak, strong and electromagnetic couplings fail to unify at high energy. In a SUSY theory, the running of the gauge couplings are modified, and precise high-energy unification of the gauge couplings is achieved:



The modified running also provides a natural mechanism for radiative EW symmetry breaking;

- *Dark matter.* TeV-scale SUSY (augmented with a discrete symmetry) typically provides a candidate dark matter particle at a mass scale consistent with thermal relic abundance calculations;
- *Other technical motivations.* SUSY is also motivated by solutions to several theoretical problems, for generally providing many desirable mathematical properties, and for ensuring sensible behavior at high energies. Supersymmetric QFT is often much easier to analyze, as many more problems become mathematically tractable.

When SUSY is imposed as a local symmetry, Einstein's theory of general relativity is included automatically, and the result is said to be a theory of *Supergravity*. It is also a necessary feature of the most popular candidate for a theory of everything, *Superstring theory*, and a SUSY theory could explain the issue of cosmological inflation.

Another theoretically appealing property of SUSY is that it offers the only "loophole" to the *Coleman-Mandula theorem*, which prohibits spacetime and internal symmetries from being combined in any non-trivial way, for QFT like the SM with very general assumptions. The *Haag-Lopuszański-Sohnius theorem* demonstrates that SUSY is the only way spacetime and internal symmetries can be combined consistently.

### 34.3 Extension of Symmetry Groups

One reason that physicists explored SUSY is because it offers an extension to the more familiar symmetries of QFT. These symmetries are grouped into the Poincaré group and internal symmetries.

The *Coleman-Mandula theorem* showed that under certain assumptions, the symmetries of the  $S$ -matrix must be a direct product of the Poincaré group with a compact internal symmetry group, or (if there is not any mass gap) the conformal group with a compact internal symmetry group.

In 1971 Golfand and Likhtman were the first to show that the Poincaré algebra can be extended through introduction of four anticommuting spinor generators (in four dimensions), which later became known as *supercharges*.

In 1975 the Haag-Łopuszański-Sohnius theorem analyzed all possible superalgebras in the general form, including those with an extended number of the supergenerators and central charges. This extended super-Poincaré algebra paved the way for obtaining a very large and important class of SUSY field theories.

Traditional symmetries of physics are generated by objects that transform by the tensor representations of the Poincaré group and internal symmetries. SUSYs, however, are generated by objects that transform by the spin representations. According to the spin-statistics theorem, bosonic fields commute while fermionic fields anticommute. Combining the two kinds of fields into a single algebra requires the introduction of a  $Z_2$ -grading under which the bosons are the even elements and the fermions are the odd elements. Such an algebra is called a *Lie superalgebra*.

The simplest SUSY extension of the Poincaré algebra is the Super-Poincaré algebra. Expressed in terms of two Weyl spinors, has the following anti-commutation relation:

$$\{Q_\alpha, \bar{Q}_{\dot{\beta}}\} = 2(\sigma^\mu)_{\alpha\dot{\beta}} P_\mu \quad (34.1)$$

and all other anti-commutation relations between the  $Q$ s and commutation relations between the  $Q$ s and  $P$ s vanish. In the above expression  $P_\mu = -i\partial_\mu$  are the generators of translation and  $\sigma^\mu$  are the Pauli matrices.

There are representations of a Lie superalgebra that are analogous to representations of a Lie algebra. Each Lie algebra has an associated Lie group

and a Lie superalgebra can sometimes be extended into representations of a Lie supergroup.

## 34.4 Applications

The mathematical structure of SUSY (graded Lie superalgebras) has subsequently been applied successfully to various topics of physics, ranging from particle and nuclear physics, critical phenomena, quantum mechanics, statistical physics to cosmology. It remains a vital part of many proposed physical theories.

### 34.4.1 The Supersymmetric SM

Incorporating SUSY into the SM requires doubling the number of particles since there is no way that any of the particles in the SM can be superpartners of each other. With the addition of new particles, there are many possible new interactions.

The simplest possible supersymmetric model consistent with the SM is the *Minimal Supersymmetric SM* (MSSM) which can include the necessary additional new particles that are able to be superpartners of those in the SM.

One of the main motivations for SUSY comes from the quadratically divergent contributions to the Higgs mass squared. The quantum mechanical interactions of the Higgs boson causes a large renormalization of the Higgs mass and unless there is an accidental cancellation, the natural size of the Higgs mass is the greatest scale possible. This problem is known as the *hierarchy problem*. SUSY reduces the size of the quantum corrections by having automatic cancellations between fermionic and bosonic Higgs interactions, since they have different signs.

If SUSY is restored at the EW scale, then the Higgs mass is related to SUSY breaking which can be induced from small non-perturbative effects explaining the vastly different scales in the weak interactions and gravitational interactions.

In many SUSY models there is a heavy stable particle (such as neutralino) which could serve as a *Weakly Interacting Massive Particle* (WIMP) dark

matter candidate. The existence of a SUSY dark matter candidate is related closely to  $R$ -parity (see below).

The standard paradigm for incorporating SUSY into a realistic theory is to have the underlying dynamics of the theory be supersymmetric, but the ground state of the theory does not respect the symmetry and SUSY is broken spontaneously.

The SUSY break cannot be done permanently by the particles of the MSSM. This means that there is a new sector of the theory that is responsible for the breaking. The only constraint on this new sector is that it must break SUSY permanently and must give superparticles TeV scale masses.

There are many models that can break SUSY and most of their details do not matter. In order to parameterize the relevant features of SUSY breaking, arbitrary soft SUSY breaking terms are added to the theory which temporarily break SUSY explicitly but could never arise from a complete theory of SUSY breaking.

One piece of evidence for MSSM is *gauge coupling unification*. The renormalization group evolution of the three gauge coupling constants of the SM is somewhat sensitive to the present particle content of the theory. These coupling constants do not quite meet together at a common energy scale if we run the renormalization group using the SM. With the addition of minimal SUSY joint convergence of the coupling constants is projected at approximately  $10^{16}$  GeV.

### 34.4.2 Supersymmetric Quantum Mechanics

SUSY quantum mechanics adds the SUSY superalgebra to quantum mechanics as opposed to QFT. This model often becomes relevant when studying the dynamics of supersymmetric solitons, and due to the simplified nature of having fields which are only functions of time (rather than space-time), a great deal of progress has been made in this subject and it is now studied in its own right.

SUSY quantum mechanics involves pairs of Hamiltonians which share a particular mathematical relationship, which are called *partner Hamiltonians* (the potential energy terms which occur in the Hamiltonians are then known as partner potentials). An introductory theorem shows that for every

eigenstate of one Hamiltonian, its partner Hamiltonian has a corresponding eigenstate with the same energy.

This fact can be exploited to deduce many properties of the eigenstate spectrum. It is analogous to the original description of SUSY, which referred to bosons and fermions. We can imagine a "bosonic Hamiltonian", whose eigenstates are the various bosons of our theory. The SUSY partner of this Hamiltonian would be "fermionic", and its eigenstates would be the theory's fermions. Each boson would have a fermionic partner of equal energy.

### 34.4.3 SUSY in Condensed Matter Physics

SUSY concepts have provided useful extensions to the WKB approximation.

Additionally, SUSY has been applied to *disorder averaged systems* both quantum and non-quantum (through statistical mechanics), an example of a non-quantum theory is the Fokker-Planck equation. The SUSY in all these systems arises from the fact that one is modelling one particle and as such the "statistics" don't matter.

The use of the SUSY method provides a mathematical rigorous alternative to the replica trick, but only in non-interacting systems, which attempts to address the so-called *problem of the denominator* under disorder averaging.

### 34.4.4 SUSY in Optics

Integrated optics was recently found to provide a fertile ground on which certain ramifications of SUSY can be explored in readily-accessible laboratory settings.

Using the mathematical structure of the SUSY Schrödinger equation and the wave equation governing the evolution of light in one-dimensional settings, one may interpret the refractive index distribution of a structure as a potential landscape in which optical wave packets propagate. Then a new class of functional optical structures with possible applications in phase matching, mode conversion and space-division multiplexing becomes possible.

SUSY transformations have been also proposed as a way to address inverse scattering problems in optics and as a one-dimensional transformation optics.

### 34.4.5 SUSY in Dynamical Systems

All stochastic (partial) differential equations, the models for all types of continuous time dynamical systems, possess so called *topological SUSY*. In the operator representation of stochastic evolution, the topological SUSY is the exterior derivative, which is commutative with the stochastic evolution operator defined as the stochastically averaged pullback induced on differential forms by diffeomorphisms of the phase space.

The topological sector of the so-emerging SUSY theory of stochastic dynamics can be recognized as the Witten-type topological field theory.

The meaning of the topological SUSY in dynamical systems is the preservation of the phase space continuity-infinately close points will remain close during continuous time evolution, even in the presence of noise.

When the topological SUSY is broken spontaneously, this property is violated in the limit of the infinitely long temporal evolution and the model can be said to exhibit (the stochastic generalization of) the *butterfly effect*.

From a more general perspective, spontaneous breakdown of the topological SUSY is the theoretical essence of the ubiquitous dynamical phenomenon known as chaos, turbulence, self-organized criticality etc. The Goldstone theorem explains the associated emergence of the long-range dynamical behavior that manifests itself as  $1/f$  noise, butterfly effect, and the scale-free statistics of sudden (instantonic) processes (e.g. earthquakes, neuro-avalanches, solar flares etc.) known as the *Zipf's law* and the *Richter scale*.

### 34.4.6 SUSY in Mathematics

SUSY is also sometimes studied mathematically for its intrinsic properties. This is because it describes complex fields satisfying a property known as *holomorphy*, which allows holomorphic quantities to be exactly computed. This makes SUSY models useful "toy models" of more realistic theories. A prime example of this has been the demonstration of *S-duality* in four-dimensional gauge theories that interchanges particles and monopoles.

Note also that the proof of the Atiyah-Singer *index theorem* is much simplified by the use of SUSY quantum mechanics.

### 34.4.7 SUSY in Quantum Gravity

SUSY is part of *superstring theory*, a string theory of quantum gravity, although it could be a component of other quantum gravity theories as well, such as *loop quantum gravity*. For superstring theory to be consistent, SUSY seems to be required (although it may be a strongly broken).

If experimental evidence confirms SUSY in the form of supersymmetric particles, such as the neutralino that is often believed to be the lightest superpartner, some physicists believe this would be a major boost to superstring theory. If major particle physics experiments fail to detect SUSY partners, many versions of superstring theory which had predicted certain low mass superpartners to existing particles may need to be significantly revised.

## 34.5 Extended SUSY

It is possible to have more than one kind of SUSY transformation. Theories with more than one SUSY transformation are known as *extended supersymmetric theories*.

The more SUSY a theory has, the more constrained are the field content and interactions. Typically the number of copies of a SUSY is a power of 2, i.e. 1, 2, 4, 8. In four dimensions, a spinor has four degrees of freedom and thus the minimal number of SUSY generators is four. Having eight copies of SUSY means that there are maximum 32 SUSY generators in four dimensions. Theories with 32 supersymmetries automatically have a graviton. Theories with more than 32 SUSY generators automatically have massless fields with spin greater than 2, for which it is not known how to define interaction.

For four dimensions there are the following theories, with the corresponding multiplets (*CPT* adds a copy, whenever they are not invariant under such symmetry):

$N = 1$ : Chiral multiplet  $(0, \frac{1}{2})$ , vector multiplet  $(\frac{1}{2}, 1)$ , gravitino multiplet  $(1, \frac{3}{2})$  and graviton multiplet:  $(\frac{3}{2}, 2)$ .

$N = 2$ : Hypermultiplet  $(-\frac{1}{2}, 0^2, \frac{1}{2})$ , vector multiplet  $(0, \frac{1}{2}, 1)$  and supergravity multiplet  $(1, \frac{3}{2}, 2)$ .



$N = 4$ : Vector multiplet  $\left(-1, -\frac{1}{2}^4, 0^6, \frac{1}{2}^4, 1\right)$  and supergravity multiplet  $\left(0, \frac{1}{2}^4, 1^6, \frac{3}{2}^4, 2\right)$ .

$N = 8$ : Supergravity multiplet  $\left(-2, -\frac{3}{2}^8, -1^{28}, -\frac{1}{2}^{56}, 0^{70}, \frac{1}{2}^{56}, 1^{28}, \frac{3}{2}^8, 2\right)$ .

### 34.5.1 SUSY and Extra Dimensions

It is possible to have SUSY in dimensions other than four. Because the properties of spinors change drastically between different dimensions, each dimension has its characteristic. In  $d$  dimensions, the size of spinors is approximately  $2^{d/2}$  or  $2^{(d-1)/2}$ . Since the maximum number of supersymmetries is 32, the greatest number of dimensions in which a SUSY theory can exist is eleven.

### 34.5.2 Fractional SUSY

Fractional SUSY is a generalization of the notion of SUSY in which the minimal positive amount of spin does not have to be  $1/2$  but can be an arbitrary  $1/N$  for integer value of  $N$ . Such a generalization is possible in two or less spacetime dimensions.

## 34.6 Experimental Searches

SUSY models are constrained by a variety of experiments, including measurements of low-energy observables - for example:

- The anomalous magnetic moment of the muon;
- The dark matter density measurement and its direct detection experiments;
- The particle collider experiments, including Higgs phenomenology,  $B$ -physics and direct searches for superpartners (sparticles).

Historically, the tightest limits were from direct production at colliders. The first mass limits for squarks and gluinos were made at CERN by the UA1 experiment and the UA2 experiment at the Super Proton Synchrotron. LEP later set very strong limits, which in 2006 were extended by the D0 experiment at the Tevatron.

From 2003-2015, WMAP's and Planck's dark matter density measurements have strongly constrained SUSY models, which, if they explain dark matter, have to be tuned to invoke a particular mechanism to sufficiently reduce the neutralino density.

Prior to the beginning of the LHC, in 2009 fits of available data to CMSSM and NUHM1 indicated that squarks and gluinos were most likely to have masses in the 500 to 800 GeV range, though values as high as 2.5 TeV were allowed with low probabilities. Neutralinos and sleptons were expected to be quite light, with the lightest neutralino and the lightest stau most likely to be found between 100 and 150 GeV.

The LHC found no evidences for SUSY, and, as a result, surpassed existing experimental limits from the Large Electron-Positron Collider and Tevatron and partially excluded the aforementioned expected ranges.

In 2011-12, the LHC discovered a Higgs boson with a mass of about 125 GeV, and with couplings to fermions and bosons which are consistent with the SM. The MSSM predicts that the mass of the lightest Higgs boson should not be much higher than the mass of the  $Z$  boson, and, in the absence of fine tuning (with the SUSY breaking scale on the order of 1 TeV), should not exceed 135 GeV.

The LHC result seemed problematic for the MSSM, as the value of 125 GeV is relatively large for the model and can only be achieved with large radiative loop corrections from top squarks, which many theorists had considered to be "unnatural".

### 34.6.1 SUSY Signatures on LHC

SUSY models allows for a wide variety of new and distinct signatures. Therefore, while searches for specific models are useful, it is also important to adopt a signature-based approach. Let us outline the main considerations in LHC serches.

1. *Interactions:* There are only two sources of model dependence here.
  - *R-parity violating couplings:* If  $R$ -parity is conserved, superpartners are produced in pairs, and each decays to a lighter superpartner plus SM particles. Thus any decay chain ends with the stable lightest superpartner. In the presence of  $R$ -parity violating couplings, a single superpartner can be produced. There are strong bounds on such couplings. Still, single superpartner production via a small  $R$ -parity violating coupling may be competitive because of the kinematics. Most superpartners decay through the usual  $R$ -parity conserving couplings (gauge or Yukawa), with the exception of the lightest superpartner, which can only decay through  $R$ -parity violating couplings.
  - *Squark and/or slepton flavor mixing:* For general sfermion mass matrices, the gaugino-sfermion-fermion couplings may mix different flavors. This could affect both production and decay.
  
2. *Superpartner masses:*
  - *The hierarchy between colored and non-colored superpartners:* In most mediation schemes, colored superpartners are heavier (by factors of a few to 10). Unless there is a huge hierarchy between colored and non-colored superpartners, the production of squarks and gluinos dominates at the LHC. As the hierarchy increases, the production of ewkinos and sleptons becomes more competitive.
  - *Flavor structure:* Superpartners of the same gauge charges may have generation dependent-masses. Thus for example, the left-handed up squark can have a different mass from the left-handed charm or top squarks. An arbitrary flavor structure leads to flavor changing processes which, especially when the first and second generations are involved, are stringently constrained. However, the constrained quantities involve the products of the mass splittings and the flavor mixings, so models with some degree of mass degeneracy and some alignment of fermion and sfermion mass matrices are allowed. These would affect both the production of sfermions and their decay.

The next to lightest superpartner plays a special role in determining the LHC signatures of SUSY. Both the mass spectrum and its interactions are relevant here.

- *The next to lightest superpartner lifetime:* The lightest superpartner is stable if  $R$ -parity is conserved. In the presence of  $R$ -parity violating couplings, the lightest superpartner can decay to SM particles, but its lifetime may be long because the sizes of these couplings are constrained. Finally, the lightest superpartner may have only very weak couplings to the SM, as is the case of the gravitino. Superpartners produced at the LHC will decay to the lightest superpartner charged under the SM, which is called the next to lightest superpartner, which in turn decays to the lightest superpartner. Clearly, since the latter is only weakly coupled to the SM, the next to lightest superpartner can be long lived. Different models span the whole range from prompt next to lightest superpartner decays to next to lightest superpartners which are long-lived on detector scales.
- *The next to lightest superpartner charge:* The next to lightest superpartner can be either neutral (e.g. a neutralino or sneutrino), charged (e.g. a slepton), or even colored (e.g. the gluino). Naturally, the precise identity of the next to lightest superpartner plays a key role in determining the signatures of SUSY at the LHC. The signatures of a spectrum with a pure bino are very different from the signatures of a spectrum with a pure Higgsino, even though both are neutral. If the next to lightest superpartner is neutral and long lived, superpartner production is accompanied by missing energy. If it is charged and long lived, it behaves like a heavy muon, and  $dE/dx$  and time-of-flight measurements must be used to distinguish it from a muon. If it is colored and long lived, it will hadronize in the detector, and subsequently either stop in the detector or traverse the entire detector. If it decays inside the detector, its charge and lifetime determine the specific signature, ranging from a disappearing track to a displaced vertex, with or without missing energy.

Thus, SUSY has motivated a wide variety of ingenious approaches for searching for new physics. Whether SUSY is there or not, this net of searches will hopefully lead to new discoveries!

**Exercise 34.1:** What is SUSY and what it is useful for?

**Exercise 34.2:** How SUSY can solve the naturalness and hierarchy problems?

**Exercise 34.3:** Explain what is a graded algebra?

## Chapter 35

### Main Concepts

So far there is no experimental evidence for SUSY and may thus be somewhat surprising the amount of effort that has been invested in SUSY, in both theory and experiment. In this respect, SUSY is not different from any other "new physics" scenario, since there is no experimental evidence for any underlying theory of EW symmetry breaking, which would give rise to the (fundamental scalar) Higgs mechanism as an effective description. However, there is of course experimental evidence for physics beyond the SM: dark matter, baryon asymmetry, neutrino masses, etc.

SUSY is popular because it is a very beautiful idea and it is conceptually different from anything we know in Nature. It is a symmetry that relates particles of different spins – bosons and fermions. In fact, we are now in a very special era from the point of view of spin, we have a spin-0 fundamental particle, Higgs boson. It would be satisfying to have some unified understanding of the spins we observe. SUSY would be a step in this direction, given the SM fermions, it predicts spin-0 particles.

Beyond this purely theoretical motivation, the fact that the Higgs is a scalar poses a more concrete (yet purely theoretical) puzzle. Scalar fields (unlike vector bosons or fermions) have quadratic divergences. This leads to the fine-tuning problem and SUSY removes these divergences. We will see that in some sense *SUSY makes a scalar behave like a fermion*.

SUSY is not a specific model, there is a wide variety of SUSY extensions of the SM. These involve different superpartner spectra, and therefore different

experimental signatures. In thinking about these, it was developed a whole toolbox for collider searches, including different triggers and analyses. In particular, SUSY supplies many concrete examples with new scalars (same charges as SM fermions), new fermions (same charges as SM gauge bosons), potentially leading to missing energy, displaced vertices, long lived charged particles, or disappearing tracks, to name just some of the possible signatures. For discovery, spin is a secondary consideration. So even if we are misguided in thinking about SUSY, and Nature is not supersymmetric, the work invested in SUSY searches may help us discover something else.

We will try to de-mystify SUSY and understand the following questions about it:

- In what sense is it a space-time symmetry (extending translations, rotations and boosts)?
- Why does it remove UV divergences (thus solving the fine-tuning, or Naturalness problem)?
- Why do we care about it even though it is clearly broken?
- Why is the gravitino relevant for LHC experiments?

## 35.1 Spacetime Symmetry

The symmetry we are most familiar with is *Poincaré symmetry*. As we know it contains:

- *Translations*:  $x^\mu \rightarrow x^\mu + a^\mu$  (generators:  $P^\mu$ );
- *Lorentz transformations*:  $x^\mu \rightarrow x^\mu + w^\mu{}_\nu x^\nu$ , where  $w^{\mu\nu}$  is antisymmetric (generators:  $J^{\mu\nu}$ ).

Throughout we will only consider global, infinitesimal transformations, so  $a^\mu$  and  $w^{\mu\nu}$  are small, coordinate-independent numbers. These transformations contain rotations. For a rotation around the axis  $x^k$  with angle  $\theta^k$  we have

$$w^{ij} = \epsilon^{ijk} \theta_k . \tag{35.1}$$

For example, for rotations around  $z$ :

$$t \rightarrow t, \quad x \rightarrow x - \theta y, \quad y \rightarrow y + \theta x, \quad z \rightarrow z. \quad (35.2)$$

The Poincaré transformations also contain boosts. For a boost along the axis  $x^k$  with the speed  $\beta^k$ :

$$-w^{0k} = w^{k0} = \beta^k. \quad (35.3)$$

Thus, for a boost along  $z$  we have,

$$t \rightarrow t + \beta z, \quad x \rightarrow x, \quad y \rightarrow y, \quad z \rightarrow z + \beta t. \quad (35.4)$$

The Poincaré algebra is:

$$\begin{aligned} [P^\mu, P^\nu] &= 0, & [P^\mu, J^{\rho\sigma}] &= 0, \\ [J^{\mu\nu}, J^{\rho\sigma}] &= i(g^{\nu\rho}J^{\mu\sigma} - g^{\mu\rho}J^{\nu\sigma} - g^{\nu\sigma}J^{\mu\rho} + g^{\mu\sigma}J^{\nu\rho}). \end{aligned} \quad (35.5)$$

Here  $P^\mu$  are the momenta – the generators of translations,  $J^{\mu\nu}$  contain the angular momenta, which generate rotations (when  $\mu, \nu = 1, 2, 3$ ) and the generators of boosts (when  $\mu = 0$  and  $\nu = 1, 2, 3$ ).

Let us recall where all this transformations is coming from. One can "discover" all the above in a simple field theory, namely a field theory of a single free complex scalar field. The Lagrangian is,

$$\mathcal{L} = \partial^\mu \phi^* \partial_\mu \phi - m^2 |\phi|^2. \quad (35.6)$$

What is a symmetry? It is a transformation of the fields which leaves the equations of motion invariant. The equations of motion follow from the action, so this tells us that the action is invariant under the symmetry transformation. Since the action is the integral of the Lagrangian, it follows that the Lagrangian can change by a total derivative,

$$\mathcal{L} \rightarrow \mathcal{L} + \alpha \partial_\mu \mathcal{J}^\mu, \quad (35.7)$$

where  $\alpha$  is the (small) parameter of the transformation.

What is the symmetry of our toy theory? First, there is a  $U(1)$  symmetry:

$$\phi(x) \rightarrow e^{i\alpha} \phi(x). \quad (35.8)$$

Under this transformation,  $\mathcal{L}$  is invariant. This is an example of an internal symmetry, that is, a symmetry which is not a space-time symmetry (it does not do anything to the coordinates). But our toy theory has also spacetime symmetries: Translations,

$$\begin{aligned} x^\mu &\rightarrow x^\mu + a^\mu , \\ \phi(x) &\rightarrow \phi(x - a) = \phi(x) - a^\mu \partial_\mu \phi(x) , \end{aligned} \quad (35.9)$$

or

$$\delta_a \phi(x) = a^\mu \partial_\mu \phi(x) , \quad (35.10)$$

and Lorentz transformations,

$$\begin{aligned} x^\mu &\rightarrow x^\mu + w^{\mu\nu} x_\nu , \\ \phi(x^\mu) &\rightarrow \phi(x^\mu - w^{\mu\nu} x_\nu) , \end{aligned} \quad (35.11)$$

giving

$$\delta_w \phi(x) = w^{\mu\nu} x_\mu \partial_\nu \phi(x) = \frac{1}{2} w^{\mu\nu} (x_\mu \partial_\nu - x_\nu \partial_\mu) \phi(x) . \quad (35.12)$$

The Lagrangian only changes by a total derivative under these, so the action is invariant.

Now let us recall how the algebra arises. Consider performing two translations. First we do a translation with  $x^\mu \rightarrow x^\mu + a^\mu$ . Then we do a translation with  $x^\mu \rightarrow x^\mu + b^\mu$ . Alternatively, we could first perform the translation with  $b^\mu$  and then the one with  $a^\mu$ . Obviously, this should not make any difference. Mathematically, this translates to the fact that the commutator of two translations vanishes. Indeed,

$$[\delta_a, \delta_b] \phi \equiv \delta_a(\delta_b \phi) - \delta_b(\delta_a \phi) = 0 . \quad (35.13)$$

However, with rotations and boosts, the order does matter. Consider the commutator of two Lorentz transformations with parameters  $w^{\mu\nu}$  and  $\lambda^{\rho\sigma}$ :

$$[\delta_{w^{\mu\nu}}, \delta_{\lambda^{\rho\sigma}}] \phi = i w_{\mu\nu} \lambda_{\rho\sigma} \cdot i [g^{\nu\rho} (x_\mu \partial_\sigma - x_\mu \partial_\sigma) + \text{permutations}] . \quad (35.14)$$

So we derived the algebra of spacetime transformations (the Poincaré algebra) in this toy example. Now let us do the same in a SUSY theory.



## 35.2 A Simple SUSY Model

Our example will be a free theory with one massive (Dirac) fermion of mass  $m$ , which we will denote by  $\psi(x)$ , and two complex scalars of the same mass  $m$ , which we will denote by  $\phi_+(x)$  and  $\phi_-(x)$ . The Lagrangian is,

$$\mathcal{L} = \partial^\mu \phi_+^* \partial_\mu \phi_+ - m^2 |\phi_+|^2 + \partial^\mu \phi_-^* \partial_\mu \phi_- - m^2 |\phi_-|^2 + \bar{\psi}(i\rlap{\not{\partial}} - m)\psi, \quad (35.15)$$

where  $\rlap{\not{\partial}} = \gamma^\nu \partial_\nu$ . The labels  $+$  and  $-$  here are just names, we will see the reason for this choice soon. This is not the most minimal SUSY field theory – ‘half of it’ is a 2-component (Weyl) fermion plus one complex scalar. But Dirac spinors are more familiar, so we start with this example.

Just as for the example in the previous section, this theory has spacetime symmetry, including translations, rotations and boosts. The only difference is that  $\psi(x)$  itself is a spinor, so it transforms nontrivially,

$$\psi(x) \rightarrow \psi'(x'). \quad (35.16)$$

Actually, the left-handed ( $L$ ) and right-handed ( $R$ ) parts of the spinor transform differently under Lorentz transformations. Write

$$\begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}, \quad (35.17)$$

where  $\psi_L$  and  $\psi_R$  are 2-component spinors. Then under Lorentz transformations (rotations by  $\theta^i$  and boosts by  $\beta^i$ ),

$$\begin{aligned} \psi_L &\rightarrow \psi'_L = \left(1 - i\theta^i \frac{\sigma^i}{2} - \beta^i \frac{\sigma^i}{2}\right) \psi_L, \\ \psi_R &\rightarrow \psi'_R = \left(1 - i\theta^i \frac{\sigma^i}{2} + \beta^i \frac{\sigma^i}{2}\right) \psi_R, \end{aligned} \quad (35.18)$$

where  $\sigma^i$  are Pauli matrices. So it will be useful to write everything in terms of 2-component spinors.

Recall that we can write any right-handed spinor in terms of a left-handed one:

$$\psi_R = -\varepsilon \chi_L^*, \quad (35.19)$$

where

$$\varepsilon \equiv -i\sigma^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (35.20)$$

We can then write our Dirac spinor in terms of two left-handed spinors  $\psi_+$  and  $\psi_-$ :

$$\psi_L = \psi_+ , \quad \psi_R = -\varepsilon\psi_-^* , \quad (35.21)$$

so that,

$$\psi = \begin{pmatrix} \psi_+ \\ -\varepsilon\psi_-^* \end{pmatrix} . \quad (35.22)$$

Let us write the Lagrangian in terms of these 2-component spinors,

$$\begin{aligned} \mathcal{L} = & \partial^\mu \phi_+^* \partial_\mu \phi_+ + \psi_+^\dagger i\bar{\sigma}^\mu \partial_\mu \psi_+ + \partial^\mu \phi_-^* \partial_\mu \phi_- + \psi_-^\dagger i\bar{\sigma}^\mu \partial_\mu \psi_- - \\ & - m^2 |\phi_-|^2 - m^2 |\phi_+|^2 - m(\psi_+^T \varepsilon \psi_- + \text{hc}) . \end{aligned} \quad (35.23)$$

All we have done so far is to re-discover spacetime symmetry in this simple field theory. Now comes the big question: *can this spacetime symmetry be extended?* The answer is yes: there is more symmetry hiding in our theory!

Take a constant (anti-commuting) 2-component left-spinor  $\xi$  and consider the following transformations,

$$\begin{aligned} \delta_\xi \phi_+ &= \sqrt{2} \xi^T \varepsilon \psi_+ , \\ \delta_\xi \psi_+ &= \sqrt{2} i \sigma^\mu \varepsilon \xi^* \partial_\mu \phi_+ - m \xi \phi_-^* , \end{aligned} \quad (35.24)$$

and similarly for  $+ \rightarrow -$ . We can see that this symmetry transformations take a boson into a fermion and vice versa – *this is SUSY*.

As an (important) aside, we note that the symmetry separately relates  $\phi_-$  to  $\psi_-$  and  $\phi_+$  to  $\psi_+$ . Thus, if  $m = 0$ , the two halves of the theory decouple and each one is symmetric separately. Therefore, as mentioned above, this theory is not the most minimal SUSY theory, but half of it. This is very handy to implement SUSY in the SM, because the SM is a chiral theory.

Is the symmetry we found indeed an extension of Poincaré? It is surely a spacetime symmetry since it takes a fermion into a boson (the transformation parameters carry spinor indices). Furthermore, let us consider the algebra. Take the commutator of two new transformations with parameters  $\xi$  and  $\eta$ :

$$[\delta_\xi, \delta_\eta] \phi_L = a^\mu \partial_\mu \phi_L , \quad (35.25)$$

where

$$a^\mu = 2i \left( \xi^\dagger \bar{\sigma}^\mu \eta - \eta^\dagger \bar{\sigma}^\mu \xi \right) . \quad (35.26)$$

This is a translation! We see that the commutator of two new transformations gives a translation. So indeed, the new symmetry is an extension of the 'usual' spacetime symmetry.

There are a couple of features of this simple example that are worth stressing because they hold quite generally:

- If the bosons and fermions had different masses, we would not have this symmetry. That is, the theory would not be supersymmetric;
- On-shell we have  $2 + 2 = 4$  fermions and  $2 + 2 = 4$  bosons, i.e. equal numbers of fermionic and bosonic degrees of freedom (off shell, the bosons are the same, but the fermions have  $2 \times 4$ ).

Let us summarize: our simple theory is supersymmetric (we have an extension of spacetime symmetry that involves anti-commuting generators) and the SUSY transformations relate bosons and fermions.

### 35.3 The Vacuum Energy

In general, global symmetries lead to Noether currents. For each global symmetry there is a current  $j^\mu$  (with  $\partial_\mu j^\mu = 0$ ), so that there is a conserved charge:

$$Q = \int d^3x j^0(x) . \quad (\dot{Q} = 0) \quad (35.27)$$

We know that for translations in time the conserved charge is the Hamiltonian  $H$ .

Thus, what we found above means that the anti-commutator of two SUSY transformations (35.25) gives the Hamiltonian. Schematically,

$$\{G_{SUSY}, G_{SUSY}\} \propto H , \quad (35.28)$$

where  $G_{SUSY}$  stands for the generator of a SUSY transformation. Now consider the vacuum expectation value (VEV) of this relation,

$$\langle 0 | \{G_{SUSY}, G_{SUSY}\} | 0 \rangle \propto \langle 0 | H | 0 \rangle . \quad (35.29)$$

If SUSY is unbroken, the ground state is supersymmetric and therefore is annihilated by the SUSY generator,

$$G_{SUSY}|0\rangle = 0 . \quad (35.30)$$

Using (35.29) we find

$$\langle 0|H|0\rangle = 0 , \quad (35.31)$$

i.e. in a SUSY theory the ground state energy is zero!

As we have already mentioned, one of the chief motivations for SUSY is the fine-tuning problem, that is, the fact that SUSY removes the quadratic divergence in the Higgs mass. Here you can already see the power of SUSY in removing UV divergences (the vacuum energy usually diverges).

This reminds the situation with an infinite constant in the energy of the harmonic oscillator, which in quantum mechanics is just set to zero by choosing the zero of the energy.

Now we see that SUSY completely removes this divergence: in a SUSY theory, the ground state energy is zero. This gives us hope that SUSY can help with other UV divergences.

The next worst divergence you can have in QFT is a quadratic divergence, which is show up in the mass-squared of scalar fields,

$$\delta m^2 \propto \Lambda^2 , \quad (35.32)$$

where  $\Lambda$  is the cutoff. This is why we are worried about fine tuning in the Higgs mass.

You could ask the question: *why nobody ever worries about the electron mass*, it too is much smaller than the Planck scale. The reason this is not a problem, is that fermion masses have no quadratic divergences, only logarithmic ones. This is a very important result so we will show it in three ways.

1. Consider a fermion Lagrangian with a mass term  $m_0$ ,

$$\mathcal{L} = \bar{\psi}(i\cancel{\partial} - m_0)\psi = \bar{\psi}(i\cancel{\partial})\psi - m_0(\psi_L^\dagger\psi_R + \psi_R^\dagger\psi_L) . \quad (35.33)$$

Note that the mass term is the only term that couples  $\psi_L$  and  $\psi_R$ . So if  $m_0 = 0$ , then the fields  $\psi_L$  and  $\psi_R$  don't talk to each other and a

mass term ( $L - R$  coupling) is never generated. Therefore, even if we include quantum corrections, we have

$$\delta m \propto m_0 , \quad (35.34)$$

where  $m$  is the full, physical mass, including quantum corrections. We see that with  $m_0 = 0$  we have two different species:  $\psi_L$ , call it, say, a 'blue' fermion, and  $\psi_R$ , a 'red' fermion, and they do not interact at all.

2. Consider  $m_0 \neq 0$ . Take a left-fermion (spin along  $\vec{p}$ ). This is our 'blue' fermion. If our speed is greater than the fermion's momentum,  $\vec{p} \rightarrow -\vec{p}$ , but the spin stays the same. Thus the fermion helicity (which is the projection of the spin along the direction of motion) changes and 'left' becomes 'right', i.e. the 'blue' fermion turns into a 'red' fermion (we know that helicity is not a good quantum number for a massive fermion). But if  $m_0 = 0$ , the 'blue' fermion is massless and it travels at the speed of light – we can never run fast enough. The 'blue' fermion does not change into a 'red' fermion. Thus, left and right are distinct in this case, and the 'blue' fermion and the 'red' fermion are decoupled.

We thus learn that any correction to the bare mass  $m_0$  must be proportional to  $m_0$ ,

$$\delta m \equiv m - m_0 \propto m_0 . \quad (35.35)$$

How can the UV cutoff  $\Lambda$  enter? On dimensional grounds,

$$\delta m = 0 \cdot \Lambda + m_0 \log \frac{m_0}{\Lambda} \propto m_0 \log \frac{m_0}{\Lambda} . \quad (35.36)$$

Indeed, there is no quadratic divergence in the fermion mass. The worst divergence that can appear is a logarithmic divergence. This is why no one ever worries about fine-tuning in the electron mass.

3. Again, the question we are asking is: *why is there no quadratic divergence in the fermion mass?* We will now see this using a global symmetry – the chiral symmetry. Let's consider the fermion Lagrangian again,

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m_0(\psi_L^\dagger\psi_R + \psi_R^\dagger\psi_L) . \quad (35.37)$$

If  $m_0 = 0$ , we have two independent  $U(1)$  symmetries,  $U(1)_L \times U(1)_R$ . This symmetry forbids the mass term. We again conclude (35.36).

We saw that SUSY implies that the boson mass equals the fermion mass. We also saw that chiral symmetry implies that there is no quadratic divergence

in the fermion mass. Putting these together we conclude: *in a SUSY theory there is no quadratic divergence in the boson mass.* This is how SUSY solves the fine-tuning problem.

But there is more that we can learn just based on dimensional analysis. We know there is no SUSY in Nature. We know for example that there is no spin-0 particle whose mass equals the electron mass. So why should we care about SUSY? The reason is that SUSY is so powerful that even when it is broken by mass terms, the quadratic divergence does not reappear! All we need in order to see this is dimensional analysis.

Suppose we take a SUSY theory and change the scalar mass (squared, since for a scalar field the physical parameter is the mass-squared that appears in Lagrangians),

$$m_{0\text{scalar}}^2 = m_{0\text{fermion}}^2 + \tilde{m}^2, \quad (35.38)$$

where  $\tilde{m}^2$  is some constant. Will there be a quadratic divergence in the scalar mass?

$$\delta m_{\text{scalar}}^2 \sim \Lambda^2 + m_{0\text{scalar}}^2 \log \frac{m_{0\text{scalar}}^2}{\Lambda^2} ? \quad (35.39)$$

No. For  $\tilde{m}^2 = 0$ , SUSY is restored, and therefore there should not be a quadratic divergence. So the  $\Lambda^2$  term (which is the quadratic divergence) must be proportional to  $\tilde{m}^2$ . But there is nothing we can write in perturbation theory that would have the correct dimension. We conclude that, if SUSY is broken by

$$m_{0\text{scalar}}^2 \neq m_{0\text{fermion}}^2, \quad (35.40)$$

*the scalar mass-squared has only log divergences.* In other words, the SUSY breaking (given by the fact that the scalar mass is different from the fermion mass) does not spoil the cancellation of the quadratic divergence.

This type of breaking is called *soft* SUSY breaking. This is what we have in the *Minimal Supersymmetric Standard Model* (MSSM).

Parenthetically, we note that one can also have *hard* SUSY breaking. Take a SUSY theory, and change some dimensionless number, e.g. the coupling of the boson compared to the coupling of the fermion. This will reintroduce the quadratic divergences.

We derived all these results based on dimensional analysis. Now let us see them concretely. To get something non-trivial we must add interactions. Let us go back to the simple toy model (35.23). Our two fermions look like the

two pieces of an electron or a quark. For example, you can think of  $\psi_-$  as the SM  $SU(2)$ -doublet quark, and of the  $\psi_+$  as the SM  $SU(2)$ -singlet quark. To get interactions we add a complex scalar  $h$  with the 'Yukawa' interaction:

$$\delta\mathcal{L} = -y h \psi_+^T \varepsilon \psi_- + \text{hc} , \tag{35.41}$$

where  $y$  is a coupling. To make a SUSY theory we also need a left fermion  $\tilde{h}$ . SUSY transformations of  $h$  and  $\tilde{h}$  are just like  $\phi_+$  and  $\psi_+$ . If  $h$  and  $\tilde{h}$  remind you of the Higgs and Higgsino that is great, but here they have nothing to do with generating mass, we are just interested in the interactions. Finally, just for simplicity, let us set  $m = 0$ .

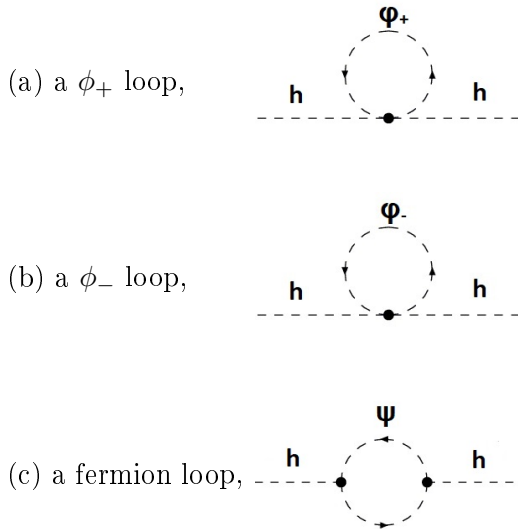
It is easy to see that if we just add this Yukawa interaction, the Lagrangian is not invariant under SUSY. So we must add more interactions,

$$\begin{aligned} \mathcal{L} = & \partial^\mu \phi_+^* \partial_\mu \phi_+ + \partial^\mu \phi_-^* \partial_\mu \phi_- + \psi_+^\dagger i \bar{\sigma}^\mu \partial_\mu \psi_+ + \psi_-^\dagger i \bar{\sigma}^\mu \partial_\mu \psi_- + \\ & + \partial^\mu h^* \partial_\mu h + \tilde{h}^\dagger i \bar{\sigma}^\mu \partial_\mu \tilde{h} + \mathcal{L}_{int} , \end{aligned} \tag{35.42}$$

with

$$\begin{aligned} \mathcal{L}_{int} = & -y \left( h \psi_+^T \varepsilon \psi_- + \phi_+ \tilde{h}^T \varepsilon \psi_- + \phi_- \tilde{h}^T \varepsilon \psi_+ + \text{hc} \right) - \\ & - |y|^2 \left( |\phi_+|^2 |\phi_-|^2 + |h|^2 |\phi_-|^2 + |h|^2 |\phi_+|^2 \right) . \end{aligned} \tag{35.43}$$

Now we have an interacting SUSY theory and are ready to consider the UV divergence in scalars mass. Consider  $\delta m_h^2$ . It gets contributions from:



To calculate the fermion loop, let us convert to Dirac fermion language,

$$y h \psi_+^T \varepsilon \psi_- + \text{hc} = y h \bar{\psi} P_L \psi + \text{hc} . \quad (35.44)$$

So the fermion loop is

$$- |y|^2 \int \frac{d^4 p}{(2\pi)^4} \text{tr} P_L \frac{i}{\not{p}} P_R \frac{i}{\not{p}} = 2 |y|^2 \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2} . \quad (35.45)$$

In the MSSM, the analog of this is the top contribution to the Higgs mass.

The boson loop is,

$$2 \times i |y|^2 \int \frac{d^4 p}{(2\pi)^4} \frac{i}{p^2} = -2 |y|^2 \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2} . \quad (35.46)$$

In the MSSM, the analog of this is the stop contribution to the Higgs mass.

Before we argued that the cancellation is not spoiled by soft SUSY breaking. Let us see this in this example. Suppose we change the  $\phi_{\pm}$  masses-squared to  $\tilde{m}_{\pm}^2$ . Indeed there is no quadratic divergence,

$$\begin{aligned} \delta m_h^2 &\propto |y|^2 \int \frac{d^4 p}{(2\pi)^4} \left[ \frac{2}{p^2} - \frac{1}{p^2 - \tilde{m}_+^2} - \frac{1}{p^2 - \tilde{m}_-^2} \right] = \\ &= |y|^2 \tilde{m}_1^2 \int \frac{d^4 p}{(2\pi)^4} \frac{1}{p^2(p^2 - \tilde{m}_+^2)} + (\tilde{m}_+^2 \rightarrow \tilde{m}_-^2) . \end{aligned} \quad (35.47)$$

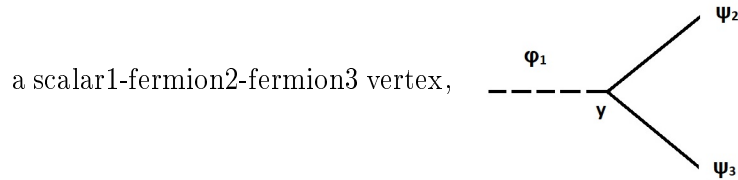
We see that when SUSY is softly broken, the scalar mass squared is log divergent, and the divergence is proportional to the SUSY breaking  $\tilde{m}^2$ . In contrast to hard SUSY breaking: if we change one of the 4-scalar couplings from  $|y|^2$ , the quadratic divergence is not cancelled.

We now know a lot of SUSY basics. Let's recap and add some language:

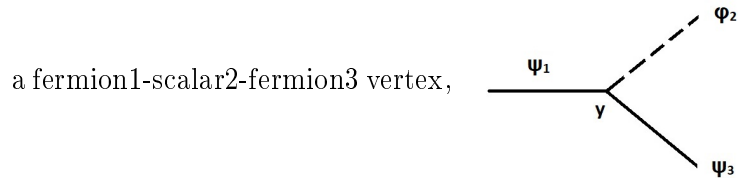
- SUSY is an extension of the Poincaré symmetry – *it is a spacetime symmetry*;
- The basic SUSY 'module' we know is a complex scalar plus a 2-component spinor of the same mass, e.g.  $(\phi_+, \psi_+)$ . These fields transform into each other under SUSY and form a representation, or a multiplet of SUSY – *the chiral supermultiplet*;
- The number of fermionic and bosonic degrees of freedom is equal to each other;



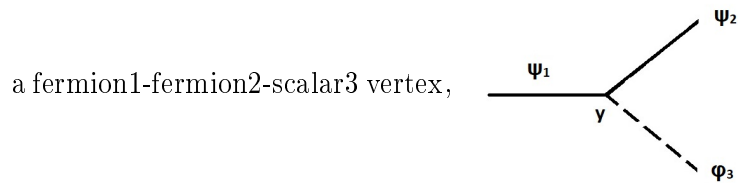
- SUSY dictates not just the field content but also the interactions. The couplings of fermions, bosons of the same supermultiplets are related. Starting from



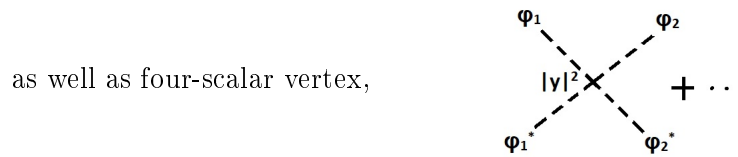
SUSY requires



and also



all with same coupling,



with the same coupling squared.

You see that the structure of supersymmetric theories is very constrained, and that as a result it is less divergent. This is the real reason theorists like SUSY, it is easier. In fact, the more SUSY, the easier it gets. There are

less divergences, more constraints, one can calculate many things, even at strong coupling. By the time you get to maximal SUSY you have a finite, scale invariant theory.

We also know a great deal about SUSY breaking, so let us summarize that too.

- With unbroken SUSY the vacuum energy is zero. Thus the vacuum energy is an order parameter for SUSY breaking, and SUSY breaking always involves a scale  $E_{\text{vac}}$ ;
- With unbroken SUSY we have only log divergences. Even in the presence of soft SUSY breaking (i.e. SUSY is broken by dimensionful quantities only), there are only log divergences;
- In contrast, hard SUSY breaking (i.e. when pure numbers, such as couplings, break SUSY) reintroduces quadratic divergences, so it is not that interesting from the point of view of the fine-tuning problem.

Let us pause and talk about language. This will be useful when we supersymmetrize the SM. Our simple example of (35.23) has two chiral supermultiplets, each contains one complex scalar and one left-handed fermion,

$$(\phi_+ \psi_+) , \quad (\phi_- \psi_-) . \quad (35.48)$$

In the SM each fermion (e.g. the top quark) comes from a fusion of two Weyl fermions: one originating from an  $SU(2)$  doublet and the other from an  $SU(2)$  singlet. These are the analogs of  $\psi_+$  and  $\psi_-$ . When we supersymmetrize the SM we must add two scalars (the stops, or top squarks) these are the analogs of  $\phi_+$  and  $\phi_-$ . One often refers to the doublet and singlet fermions as 'left-handed' and 'right-handed'.

This is bad language (remember we can always write a left handed spinor using a right-handed spinor). If we used this bad language anyway, we could call our fermions  $\psi_L$  and  $\psi_R$ , and the accompanying scalars:  $\phi_L$  and  $\phi_R$ . This is why you hear people talk about the stop-left and stop-right, or left squarks and right squarks. Of course the stops are scalars, and have no chirality, but the names just refer to their fermionic partners.

## 35.4 SUSY Breaking

If SUSY is realized in Nature it is realized as a broken symmetry. We already saw that even explicit (soft) SUSY breaking can be powerful. But the picture we had is not very satisfying: we don't want to put in the parameter  $\tilde{m}^2$  by hand. We want it to be generated by the theory itself, i.e. we want the theory to break SUSY spontaneously.

We also saw that with unbroken SUSY the vacuum energy vanishes, and the potential  $V \geq 0$ . Thus SUSY is unbroken if there are solution(s) of the equations of motions with  $V = 0$ . Recall that this followed from

$$\langle 0 | \{G_{SUSY}, G_{SUSY}\} | 0 \rangle \propto \langle 0 | H | 0 \rangle , \quad (35.49)$$

and with unbroken SUSY,

$$G_{SUSY} | 0 \rangle = 0 . \quad (35.50)$$

However, if SUSY is spontaneously broken:

$$G_{SUSY} | 0 \rangle \neq 0 , \quad (35.51)$$

and the ground state has nonzero (positive) energy!

In the SM, the only scalar is the Higgs, so the only potential is the Higgs potential. But in SUSY theories, fermions are always accompanied by scalars and any fermion interaction results in a scalar potential, e.g.

$$V(\phi_+, \phi_-, h) = |y|^2 (|\phi_+|^2 |\phi_-|^2 + |h|^2 |\phi_-|^2 + |h|^2 |\phi_+|^2) . \quad (35.52)$$

To break SUSY spontaneously all we need is to find a SUSY theory with a potential which is always above zero. So we need a scale. Classically, we can just put in scale by hand. This brings us to the simplest example of spontaneous SUSY breaking.

### 35.4.1 The O'Raifeartaigh Model

The simplest SUSY theory with chiral supermultiplets that breaks SUSY spontaneously has three chiral supermultiplets,

$$(\phi, \psi) , \quad (\phi_1, \psi_1) , \quad (\phi_2, \psi_2) , \quad (35.53)$$

and two mass parameters. We will only write the scalar potential (the kinetic terms and fermion-fermion-scalar interactions are there too, but there is nothing instructive in them at this point),

$$V = |y\phi_1^2 - f|^2 + m^2|\phi_1|^2 + |2\phi_1\phi + m\phi_2|^2 . \quad (35.54)$$

Here  $m$  is a mass,  $f$  has dimension mass<sup>2</sup>, and  $y$  is a dimensionless coupling. It is easy to see that there is no supersymmetric minimum. The first two terms cannot vanish simultaneously. SUSY is broken! Note that we need  $f \neq 0$  for that (we must push some field away from the origin) as well as  $m \neq 0$ .

Finding the ground state requires more effort. Let us assume

$$f < \frac{m^2}{2y} . \quad (35.55)$$

The ground state is at

$$\phi_1 = \phi_2 = 0 \quad (35.56)$$

with  $\phi$  arbitrary ( $\phi$  is a flat direction of the potential),

$$V_0 = |f|^2 . \quad (35.57)$$

Expanding around the VEVs, one finds the following spectrum: one massless Weyl fermion, one Dirac fermion of mass  $m$ , and several real bosons of which two are massless, two have mass  $m$ , one has mass  $\sqrt{m^2 + 2yf}$ , and one  $\sqrt{m^2 - 2yf}$ . Indeed, for  $f = 0$  SUSY is restored, and the fermions and bosons become degenerate.

Why are there massless bosons in the spectrum? Recall that  $\phi$  is arbitrary, it is a flat direction (two real degrees of freedoms). Why is there a massless Weyl fermion? Normally a global SSB implies the existence of a massless Goldstone boson (or pion). Here we have spontaneously broken SUSY, which is a 'fermionic' symmetry, so we have a massless Goldstone fermion. Since SUSY is broken spontaneously, the SUSY generator does not annihilate the vacuum,

$$G_{SUSY}|0\rangle \neq 0 . \quad (35.58)$$

Since this generator carries a spinor index, this state is a fermion state, which is precisely the Goldstone fermion (sometimes called a Goldstino).

Recall that we needed a scale, or a dimensionful parameter in order to break SUSY. Above we simply put it in by hand. But suppose we started with

no scale in the Lagrangian. Then classically SUSY would remain unbroken. This suggests that, since no scale can be generated perturbatively, if SUSY is unbroken at the tree-level, it remains unbroken to all orders in perturbation theory.

This is actually true, and it is a very powerful result. It is a consequence of the constrained structure of SUSY. So if SUSY is unbroken at tree level, it can only be broken by non-perturbative effects, with a scale that is generated dynamically, just like the QCD scale,

$$\Lambda = M_{UV} e^{-8\pi^2/bg^2} , \quad (35.59)$$

which is exponentially suppressed compared to the cutoff scale.

This type of SUSY breaking is called dynamical SUSY breaking. We will come back to this when we discuss the SM. It leads to a beautiful scenario: *the SUSY breaking scale can naturally be 16 or so orders of magnitude below the Planck scale.*

## 35.5 SUSY Lagrangians

Suppose we have the chiral supermultiplet:

$$(\phi, \psi) , \quad (35.60)$$

with  $\phi$  a complex scalar,  $\psi$  a 2-component fermion. In the real world we also have spin-1 gauge bosons,  $A_\mu^a$ , where  $a$  denotes the gauge group index. So in order to supersymmetrize the SM we also need vector supermultiplets,

$$(A_\mu^a, \lambda^a) , \quad (35.61)$$

namely, a gauge field plus a 'gaugino'. On-shell  $A_\mu^a$  has two degrees of freedom (two physical transverse polarizations), so  $\lambda^a$  is a 2-component spinor.  $A_\mu^a$  is real, so if we want to write  $\lambda^a$  as a 4-component spinor it must be a Majorana spinor (as opposed to the Dirac fermion which consists of two distinct 2-component fermions),

$$\begin{pmatrix} \lambda \\ -\varepsilon\lambda^* \end{pmatrix} . \quad (35.62)$$

Under a SUSY transformation,  $A_\mu^a$  and  $\lambda^a$  transform into each other, and we can construct SUSY Lagrangians for them as we did for the chiral supermultiplet.

We now have the gauge module (gauge field + gaugino) and the chiral module (scalar + fermion). What are the Lagrangians we can write down? With a theory of such a constrained structure, you expect to have many limitations. Indeed, all the theories we can write down are encoded by two functions, the Kähler potential ( $K$ ), which gives the kinetic and gauge interactions (as we will see, there is no freedom there, so we won't even write it down), and the superpotential ( $W$ ), which gives the non-gauge (Yukawa-like) interactions of chiral fields.

Let us start with the gauge part: after all, gauge interactions are almost all we measure.

### 35.5.1 A SUSY Gauge Theory

We want a gauge-invariant SUSY Lagrangian for

$$(A_\mu^a, \lambda^a) . \quad (35.63)$$

Gauge symmetry and SUSY determine this Lagrangian completely up to higher-dimension terms. It is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F^{a\mu\nu} F_{\mu\nu}^a + \lambda^{a\dagger} i \bar{\sigma} \cdot \mathcal{D} \lambda^a . \quad (35.64)$$

We also want to couple 'matter fields' to the gauge field. So we add our chiral modules  $(\phi_i, \psi_i)$ . There is no freedom here too, because of gauge symmetry plus SUSY,

$$\begin{aligned} \mathcal{L} = & \mathcal{L}_{\text{gauge}} + \mathcal{D}^\mu \phi_i^* \mathcal{D}_\mu \phi_i + \psi_i^\dagger i \bar{\sigma}^\mu \mathcal{D}_\mu \psi_i - \\ & - \sqrt{2} g (\phi_i^* \lambda^{aT} T^a \varepsilon \psi_i - \psi_i^\dagger \varepsilon \lambda^{a*} T^a \phi_i) - \frac{1}{2} D^a D^a , \end{aligned} \quad (35.65)$$

where

$$D^a = -g \phi_i^\dagger T^a \phi_i . \quad (35.66)$$

As in other chiral theories, SUSY dictates 'new' couplings. In non-SUSY theories we have a coupling *gauge field-fermion-fermion*. Now we also have *gaugino-fermion-scalar* and of course there is also *gauge field-scalar-scalar*.

In addition, there is a scalar potential with a 4-scalar interaction,

$$V = \frac{1}{2} D^a D^a , \quad (35.67)$$

with

$$D^a = -g\phi_i^\dagger T^a \phi_i . \quad (35.68)$$

This is a quartic scalar potential, but the quartic coupling is not arbitrary, it is the gauge coupling. This will be very important when we discuss the Higgs!

Note what happened here. Starting from a non-SUSY gauge theory with a gauge field  $A_\mu^a$  and a fermion  $\psi$ , and an interaction  $gA_\mu\psi\psi$ , when we supersymmetrize the theory, the field content is gauge field + gaugino, fermion + scalar. The interactions are  $A_\mu\phi\phi$  (nothing new,  $\phi$  is charged), but also,  $\lambda\phi\psi$  (gaugino-scalar-fermion) all with same coupling  $g$ . In addition there is a 4-scalar interaction, with coupling  $g^2$ .

We had no freedom in the process. The field content and couplings of the SUSY theory were dictated by:

- the original non-SUSY theory we started from;
- the gauge symmetry;
- SUSY.

### 35.5.2 Yukawa-like Interactions

We also want Yukawa-like interactions of just the chiral scalars and fermions ( $\phi_i, \psi_i$ ). There is a simple recipe for writing down the most general supersymmetric interaction Lagrangian. Choose an analytic function  $W(\phi_1, \dots, \phi_n)$  – the 'superpotential'. Analytic means that  $W$  is not a function of the conjugate fields (no daggers!). All the allowed interactions are given by

$$\mathcal{L}_{\text{int}} = -\frac{1}{2} \frac{\partial^2 W}{\partial \phi_i \partial \phi_j} \psi_i^T \varepsilon \psi_j + \text{hc} - \sum_i |F_i|^2 , \quad (35.69)$$

where

$$F_i^* = -\frac{\partial W}{\partial \phi_i} . \quad (35.70)$$

This Lagrangian is guaranteed to be supersymmetric!

Let us write our previous examples in this language. Start with the theory containing  $h$ ,  $\phi_+$ ,  $\phi_-$  and take

$$W = y h \phi_+ \phi_- . \quad (35.71)$$

So

$$\begin{aligned} F_h^* &= -\frac{\partial W}{\partial h} = y \phi_+ \phi_- , \\ \frac{\partial^2 W}{\partial h \partial \phi_+} &= y \phi_- , \end{aligned} \quad (35.72)$$

and similarly for the remaining fields. We indeed recover

$$\begin{aligned} \mathcal{L}_{int} = & - y \left( h \psi_+^T \varepsilon \psi_- + \phi_+ \tilde{h}^T \varepsilon \psi_- + \phi_- \tilde{h}^T \varepsilon \psi_+ + \text{hc} \right) - \\ & - |y|^2 (|\phi_+|^2 |\phi_-|^2 + |h|^2 |\phi_-|^2 + |h|^2 |\phi_+|^2) . \end{aligned} \quad (35.73)$$

### 35.5.3 $R$ -symmetry

We now know how to write the most general supersymmetric Lagrangian (with minimal SUSY). Note that the theory has a global  $U(1)$  symmetry. Under this  $U(1)$ , the gauge boson has charge 0, the gaugino has charge +1, the chiral fermion has charge 0 and the scalar  $-1$ .

Alternatively we could take the fermion to have charge  $-1$  and the scalar with charge 0. This is called a  $U(1)_R$  symmetry. It does not commute with SUSY: members of the same supermultiplet have different charges.

This symmetry (or its remanent) is crucial in LHC supersymmetry searches!

### 35.5.4 $F$ -terms and $D$ -terms

In writing SUSY Lagrangians, we are defining so call  $F$  and  $D$  terms. For each vector multiplet  $(\lambda^a, A_\mu^a)$  we have

$$D^a = -g \phi_i^\dagger T^a \phi_i \quad (\text{dim} - 2) \quad (35.74)$$

(to which all the charged scalars contribute).



For each chiral multiplet  $(\phi_i, \psi_i)$  we have

$$F_i^* = -\frac{\partial W}{\partial \phi_i}, \quad (\text{dim} - 2) \quad (35.75)$$

and the scalar potential is

$$V = F_i^* F_i + \frac{1}{2} D^a D^a. \quad (35.76)$$

With this language, we can revisit SUSY breaking. First we immediately see

$$V \geq 0. \quad (35.77)$$

Second, SUSY is broken if some  $F_i \neq 0$  or  $D^a \neq 0$ . So the  $F_i$ 's and  $D^a$ 's are order parameters for SUSY breaking.

### 35.5.5 Local SUSY and Supergravity

So far we thought about SUSY as a global symmetry of Lagrangians. Translations and Lorentz transformations are however local symmetries. The 'gauge theory' of local spacetime symmetry is gravity. We therefore have no choice: SUSY is a local symmetry too.

The theory of local (spacetime and) SUSY is called *Supergravity*. The spin-2 graviton must have a supersymmetric partner, the gravitino, which has spin-3/2. Since SUSY is broken the gravitino should get mass.

If you are only interested in collider experiments, should you care about this? Normally, the effects of gravity are suppressed by the Planck scale and we can forget about them when discussing high energy experiments. However, the gravitino mass is related to a broken local symmetry (SUSY), and, just as in the usual Higgs mechanism of EW symmetry breaking, it gets mass by 'eating' the Goldstone fermion.

Thus, a piece of the gravitino (the longitudinal piece), is some 'ordinary' field (which participates in SUSY breaking), and the gravitino couplings to matter are not entirely negligible.

Furthermore, they are dictated by the SUSY breaking. If SUSY is broken by some non-zero  $F$  term, the gravitino mass is

$$m_{3/2} \sim \frac{F}{M_{Pl}}. \quad (35.78)$$

**Exercise 35.1:** Write out commutation relations of the  $N = 1$  SUSY extension of the Poincaré algebra.

**Exercise 35.2:** Find commutation relations of the Pauli-Ljubanski vector with the generators of the  $N = 1$  SUSY Poincaré algebra and show that the second Casimir of the ordinary Poincaré group indeed commutes with the Poincaré generators but not with its SUSY extension.

**Exercise 35.3:** Derive the transformation rules of a general scalar superfield with single spinor variable under super Poincaré.

**Exercise 35.4:** Find the combination of SUSY covariant derivatives and its conjugates which leads to the Klein-Gordon operator.

**Exercise 35.5:** Prove (35.19).

**Exercise 35.6:** Derive (35.23). Show also that  $\psi_+^T \varepsilon \psi_- = \psi_-^T \varepsilon \psi_+$ , where  $\psi_{\pm}$  are any 2-component spinors.

**Exercise 35.7:** Show that the massless part of the Lagrangian is invariant under (35.24) and that the rest of the Lagrangian is invariant too if the masses of the fermion and scalars are the same.

**Exercise 35.8:** Check (35.25).

**Exercise 35.9:** Check that the Lagrangian (35.64) is supersymmetric.

**Exercise 35.10:** Check that the massive theory with  $\phi_{\pm}$  is obtained from  $W = m\phi_+\phi_-$ .

**Exercise 35.11:** Check that the O’Raifeartaigh model is obtained from  $W = \phi(y\phi_1^2 - f) + m\phi_1\phi_2$ .

# Chapter 36

## The SUSY and SM

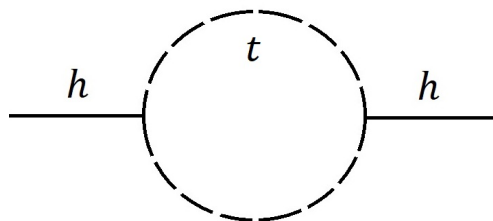
Now that we understand what SUSY is, we can supersymmetrize the SM. Let us review first the motivations for doing that.

### 36.1 Motivation

Before 2012 all fundamental particles we knew had spin 1, or spin 1/2. Now we have the Higgs with the spin 0, this is the source of the *fine-tuning*, or *naturalness problem*. Since the Higgs is spin-0, its mass is quadratically divergent

$$\delta m^2 \propto \Lambda_{UV}^2, \tag{36.1}$$

unlike fermions, whose masses are protected by the chiral symmetry as we saw, or gauge bosons, whose masses are protected by gauge symmetry. In the case of the Higgs mass, there are one loop corrections that are quadratically divergent. The dominant one is from the top quark.



This is not a practical problem. We can calculate any physical observable by including a counter term that cancels this divergent contribution. Rather, the problem is of a theoretical nature. We believe that  $\Lambda_{UV}$  is a concrete physical scale, such as the mass scale of new fields, or the scale of new strong interactions. Then at the low-scale  $\mu$ ,

$$m^2(\mu) = m^2(\Lambda_{UV}) + c\Lambda_{UV}^2 , \quad (36.2)$$

where  $m^2(\Lambda_{UV})$  determined by the full UV theory, and the constant  $c$  determined by the SM. We know the left hand side of (36.2):  $m^2 \sim 100 \text{ GeV}^2$ . So if  $\Lambda_{UV}$  is the Planck scale  $\sim 10^{18} \text{ GeV}$  we need  $m^2(\Lambda_{UV}) \sim 10^{36} \text{ GeV}^2$  and the two terms on the right hand side must be tuned to 32 orders of magnitude. Such dramatic *tunings* do not seem *natural*. In general, for a cutoff scale  $\Lambda_{UV}$ , the parameters of the two theories must be tuned to  $\text{TeV}^2/\Lambda_{UV}^2$ .

As we saw above, with SUSY (even softly broken), scalar masses-squared have only log divergences:

$$m^2(\mu) = m^2(\Lambda_{UV}) \left[ 1 + c \log \left( \frac{m^2(\Lambda_{UV})}{\Lambda_{UV}^2} \right) \right] , \quad (36.3)$$

just as for fermions! The reason is that SUSY ties the scalar mass to the fermion mass.

The way this happens is that the quadratic divergence from fermion loops is cancelled by the quadratic divergence from scalar loops. The cutoff scale then only enters in the log, and  $m^2(\Lambda_{UV})$  can be order  $(100 \text{ GeV})^2$ . This is the main motivation for SUSY extensions of the SM.

There are further motivations too. SUSY extensions of the SM often supply dark matter candidates, new sources of  $CP$  violation, etc. Finally, extending space time symmetry is theoretically appealing.

## 36.2 The Structure of the Model

So let us supersymmetrize the SM. Each gauge field of the SM will be now part of a vector supermultiplet. For the gluon we have,

$$G_\mu^a \rightarrow (\tilde{g}^a, G_\mu^a) + D^a , \quad (36.4)$$

where the physical fields are the gluon and the spin-1/2 gluino. Similarly for the  $W$  boson,

$$W_\mu^I \rightarrow (\tilde{w}^I, W_\mu^I) + D^I , \quad (36.5)$$

where the physical fields are the  $W$  and the wino, and for  $B$  boson,

$$B_\mu \rightarrow (\tilde{b}, B_\mu) + D_Y , \quad (36.6)$$

where the physical fields are the  $B$  and the bino.

Each fermion is now part of a chiral supermultiplet of the form:

$$(\phi, \psi) + F . \quad (36.7)$$

Taking all the SM fermions  $q, u^c, d^c, l, e^c$  to be left-fermions, we have:

$$q \rightarrow (\tilde{q}, q) + F_q , \quad \text{all transforming as } (3, 2)_{1/6} , \quad (36.8)$$

with the physical fields being the (doublet) quark  $q$  and a spin-0 squark  $\tilde{q}$ .

Similarly, for singlets,

$$u^c \rightarrow (\tilde{u}^c, u^c) + F_u , \quad \text{all transforming as } (\bar{3}, 1)_{-2/3} , \quad (36.9)$$

with the physical fields being the (singlet) up-quark  $u^c$  + up squark  $\tilde{u}^c$ ,

$$d^c \rightarrow (\tilde{d}^c, d^c) + F_d , \quad \text{all transforming as } (\bar{3}, 1)_{1/3} , \quad (36.10)$$

with the physical fields being the (singlet) down-quark  $d^c$  + down squark  $\tilde{d}^c$ .

For leptons,

$$l \rightarrow (\tilde{l}, l) + F_l , \quad \text{all transforming as } (1, 2)_{-1/2} , \quad (36.11)$$

with the physical fields being the (doublet) lepton  $l$  + a slepton  $\tilde{l}$ , and finally

$$e^c \rightarrow (\tilde{e}^c, e^c) + F_e , \quad \text{all transforming as } (1, 1)_1 , \quad (36.12)$$

with the physical fields being the (singlet) lepton  $e^c$  + a slepton  $\tilde{e}^c$ .

Once EW symmetry is broken the quark and lepton doublets split:

$$q = \begin{pmatrix} u \\ d \end{pmatrix} , \quad \tilde{q} = \begin{pmatrix} \tilde{u} \\ \tilde{d} \end{pmatrix} , \quad (36.13)$$

and

$$l = \begin{pmatrix} \nu \\ l \end{pmatrix}, \quad \tilde{l} = \begin{pmatrix} \tilde{\nu} \\ \tilde{l} \end{pmatrix}. \quad (36.14)$$

Now let us move on to the interactions, starting with the gauge interactions. There is nothing we have to do here. As we saw above, these interactions are completely dictated by SUSY and the gauge symmetry.

In previous chapter we wrote the Lagrangian for a general SUSY gauge theory in the form (35.65). Applying this to the SM,

$$\psi_i = q_i, u_i^c, d_i^c, l_i, e_i^c, \quad \phi_i = \tilde{q}_i, \tilde{u}_i^c, \tilde{d}_i^c, \tilde{l}_i, \tilde{e}_i^c. \quad (36.15)$$

The covariant derivatives now contain the  $SU(3)$ ,  $SU(2)$  and  $U(1)$  gauge fields,  $\lambda^a$  sums over the  $SU(3)$ ,  $SU(2)$  and  $U(1)$  gauginos

$$\lambda^a \rightarrow \tilde{g}^a, \tilde{w}^I, \tilde{b}. \quad (36.16)$$

Also there are  $D$  terms for  $SU(3)$ ,  $SU(2)$  and  $U(1)$ ,

$$D^a \rightarrow D^a, D^I, D_Y. \quad (36.17)$$

In addition there is of course the pure gauge Lagrangian, this part we don't consider now.

The Lagrangian (35.65) contains the scalar potential,

$$V = \frac{1}{2} D^a D^a + \frac{1}{2} D^I D^I + \frac{1}{2} D_Y D_Y, \quad (36.18)$$

where for  $SU(3)$ :

$$D^a = g_3 (\tilde{q}^\dagger T^a \tilde{q} - \tilde{u}^{c\dagger} T^{a*} u^c - \tilde{d}^{c\dagger} T^{a*} u^c), \quad (36.19)$$

recall  $T_3 = -T_3^*$  and we will write things in terms of the fundamental generators. Similarly for the  $SU(2)$ ,

$$D_Y = g_Y \sum_i Y_i \tilde{f}_i^\dagger \tilde{f}_i. \quad (36.20)$$

We see that we get 4-scalar interactions with the quartic couplings equal to the gauge couplings.

Again we emphasize that there was no freedom so far, and no new parameters. We also did not put in the Higgs field yet, so let us do this now. The SM Higgs is a complex scalar, so it must be part of a chiral module

$$H \rightarrow (H, \tilde{H}) + F_H, \quad \text{all transforming as } (1, 2)_{-1/2}. \quad (36.21)$$

We immediately see problems with having a single Higgs scalar, which are all related:

1. We want the Higgs to get a VEV. However, the Higgs is charged under  $SU(2)$  and  $U(1)$ , so its VEV gives rise to nonzero  $D$  terms:

$$V \sim D^I D^I + D_Y^2, \quad (36.22)$$

where

$$D^I = g_2 \langle H^\dagger \rangle T^I \langle H \rangle, \quad D_Y = g_1 \frac{1}{2} \langle H \rangle^\dagger \langle H \rangle, \quad (36.23)$$

i.e. EW SSB implies SUSY breaking!

You might think this is good, but it is not (for many reasons). For one, the non-zero  $D$ -terms would generate masses for the squarks and sleptons. Consider  $D_Y$  for example:

$$D_Y = \frac{1}{2} v^2 + \sum_i Y_i |\tilde{f}_i|^2, \quad (36.24)$$

where  $\tilde{f}$  sums over all squarks, sleptons and  $Y_i$  is their hypercharge. Recall the scalar potential  $V \sim D^2$ . Therefore some of the squarks will get negative masses-squared of order  $v^2$ .

This is a disaster:  $SU(3)$  and EW symmetries are broken at  $v$ ! The solution is to add a second Higgs scalar, with opposite charges. The two Higgs scalars can then get equal VEVs with all  $\langle D \rangle = 0$ .

2. A second problem is that  $\tilde{H}$  is a Weyl fermion. Then we will have a massless fermion around – the Higgsino. In the presence of massless fermions gauge symmetries can become anomalous, that is, the gauge symmetry can be broken at the loop level. In the SM the fermion representations and charges are such that there are no anomalies.

Before discussing the Higgs, we only added scalars to the SM (squarks and sleptons, known collectively as sfermions). These are harmless

from the point of view of anomalies. We also added gauginos. These are fermions, but they are adjoint fermions, which don't generate any anomalies (essentially because the adjoint is a real representation). In contrast, the Higgsino  $\tilde{H}$  is a massless fermion which is a doublet of  $SU(2)$  and charged under  $U(1)_Y$ . The simplest way to cancel the anomaly is to add a second Higgsino in the conjugate representation. So we must add a second Higgs field with conjugate quantum numbers.

3. When we consider interactions, we will see other reasons why we must add another Higgs. We will call the SM Higgs  $H_D$  and the new Higgs  $H_U$ . Thus,

$$H_D \rightarrow (H_D, \tilde{H}_D) + F_{HD} , \quad \text{all transforming as } (1, 2)_{-1/2} , \quad (36.25)$$

and we also add,

$$H_U \rightarrow (H_U, \tilde{H}_U) + F_{HU} , \quad \text{all transforming as } (1, 2)_{1/2} , \quad (36.26)$$

and in the limit of unbroken SUSY,

$$\langle H_U \rangle = \langle H_D \rangle . \quad (36.27)$$

In the SM we add a quartic potential for the Higgs field,

$$\lambda(H^\dagger H)^2 . \quad (36.28)$$

Here there is quartic potential built in, coming from the  $D$  terms. This potential will not necessarily give mass to the physical Higgs.

We now turn to the Yukawa couplings. In the SM we have Higgs-fermion-fermion Yukawa couplings. Consider the down-quark Yukawa first,

$$y_D H_D q^T \varepsilon d^c \quad (\text{Higgs} - \text{quark} - \text{quark}) , \quad (36.29)$$

as we saw above, with SUSY this must be accompanied by

$$y_D (\tilde{q} \tilde{H}_D^T \varepsilon d^c + \tilde{d}^c \tilde{H}_D^T \varepsilon q) \quad (\text{squark} - \text{Higgsino} - \text{quark}) , \quad (36.30)$$

all coming from the superpotential

$$W_D = y_D H_D q d^c . \quad (36.31)$$

Similarly for the lepton Yukawa,

$$\begin{aligned} W_l = y_l H_D l e^c \rightarrow \mathcal{L}_l = & y_l (H_D l^T \varepsilon e^c + \tilde{l} \tilde{H}_D^T \varepsilon e^c + \tilde{e}^c \tilde{H}_D^T \varepsilon l + \text{hc}) + \\ & + (\text{Higgs} - \text{lepton} - \text{lepton}) + \quad (36.32) \\ & + (\text{slepton} - \text{Higgsino} - \text{lepton}) . \end{aligned}$$



4. What about the up Yukawa? We need,

$$(\text{Higgs})q^T \varepsilon u^c . \quad (36.33)$$

This coupling must come from a superpotential,

$$(\text{Higgs})qu^c . \quad (36.34)$$

In the SM  $(\text{Higgs}) = H_D^\dagger$ . But the superpotential is holomorphic, no daggers are allowed. This is the fourth reason why we needed a second Higgs field with the conjugate charges,

$$\begin{aligned} W_U &= y_U H_U q u^c \rightarrow \\ \mathcal{L}_U &= y_U (H_U q^T \varepsilon u^c + \tilde{q} \tilde{H}_U^T \varepsilon u^c + \tilde{u}^c \tilde{H}_U^T \varepsilon q) + \text{hc} . \end{aligned} \quad (36.35)$$

You can now see what is going on. In some sense, holomorphy makes a scalar field "behave like a fermion". In a SUSY theory the interactions of scalar fields are controlled by the superpotential, which is holomorphic. For a fermion to get mass you need a left-right coupling. So starting from a left-fermion you need a right-fermion, or another left-fermion with the opposite charge(s). For a scalar  $\phi$  to get mass in a non-SUSY theory you don't need anything else (you can just use  $\phi^*$  to write a charge neutral mass term). Not so in a SUSY theory: because you cannot use  $\phi^*$ , you must have another scalar with the opposite charge(s), just as for fermions.

To summarize, we have two Higgs fields  $H_U$  and  $H_D$  and the SM Yukawa couplings come from the superpotential:

$$W = y_U H_U q u^c + y_D H_D q d^c + y_l H_D l e^c . \quad (36.36)$$

Note again that there was no freedom here, and no new parameter.

### 36.2.1 *R*-symmetry

Note that our SUSY Lagrangian also has a  $U(1)_R$  symmetry. Here is one possible choice of charges: gaugino ( $-1$ ), sfermions ( $1$ ), Higgsinos ( $1$ ), with all other fields, namely the SM fields, neutral. In each of the interactions, the new superpartners appear in pairs! This is important both for LHC production and for dark matter.

To recap, we wrote down the SUSY SM which contains:

- gauge bosons + (spin-1/2) gauginos;
- fermions + (spin-0) sfermions;
- 2 Higgses + 2 (spin-1/2) Higgsinos.

The interactions are all dictated by the SM interactions + SUSY. The new interactions are:

- gauge-boson – scalar – scalar;
- gauge-boson – gauge-boson – scalar – scalar;
- gaugino – sfermion – fermion;
- gauge-boson – Higgsino – Higgsino;
- 4-scalar (all gauge invariant contributions).

All the couplings are determined by the SM gauge couplings. In particular, there is a quartic Higgs coupling which is proportional to the gauge-coupling squared.

Furthermore, there is the Yukawa part, which now contains: Higgsino – fermion – sfermion, with a coupling equal to the SM Yukawa coupling.

There is now no quadratic divergence in the Higgs mass. Each quark contribution is canceled by the corresponding squark contribution. In particular the top loop is canceled by the left, right stops. Similarly, the contribution from the Higgs self-coupling (from the  $D$  term) is canceled by the Higgsinos, and each gauge boson contribution is canceled by the gaugino contribution.

But we now have, a wino degenerate with the  $W$ , a selectron degenerate with the electron, etc. SUSY must be broken. Somehow the wino, selectron, and all the new particles should get mass. It would be nice if the supersymmetrized SM broke SUSY spontaneously (after all we have lots of scalars with a complicated potential). But it does not, and so we must add more fields and interactions that break SUSY. These new fields must couple to the SM fields in order to generate masses for the superpartners.

## 36.3 SUSY Breaking

For SSB of SUSY we need a set of new fields and interactions. As a result there will be mass splittings between the bosons and fermions of the same symmetry breaking multiplet. Since there are SUSY-breaking mass-splittings among the symmetry breaking fields, their coupling will generate mass splitting between the SM fields and their superpartners, mediating the SUSY breaking to the MSSM. The mediation mechanism determines the SUSY-breaking terms in the MSSM, which in turn determine the experimental signatures of SUSY.

### 36.3.1 Breaking Terms

A Lagrangian allows any term unless a symmetry prevents it. If we assume SUSY breaking terms, in the matter sector, sfermions get mass. However, the fermions don't: they are protected by chiral symmetry. In the gauge sector, gauginos get mass. However, gauge bosons don't: they are protected by gauge symmetry. In the Higgs sector, the Higgses get mass. Higgsinos don't, they are protected by chiral symmetry. This is a problem. We would like the gauginos to get mass, so we will have to solve this problem.

In addition, there are trilinear scalar terms that can appear, such as a Higgs – squark – squark coupling, or a Higgs – slepton – slepton coupling. These are allowed by gauge symmetry, and SUSY is no longer there to forbid them. These terms are called  $A$ -terms.

The SUSY-breaking part of the MSSM Lagrangian is:

$$\begin{aligned}
 \mathcal{L}_{\text{soft}} = & -\frac{1}{2} \left[ \tilde{m}_3 \tilde{g}^T \varepsilon \tilde{g} + \tilde{m}_2 \tilde{w}^T \varepsilon \tilde{w} + \tilde{m}_1 \tilde{b}^T \varepsilon \tilde{b} \right] - \tilde{q}^* \tilde{m}_q^2 \tilde{q} - \\
 & - \tilde{u}^{c*} \tilde{m}_{uR}^2 \tilde{u}^c - \tilde{d}^{c*} \tilde{m}_{dR}^2 \tilde{d}^c - \tilde{l}^* \tilde{m}_l^2 \tilde{l} - \tilde{e}^{c*} \tilde{m}_{eR}^2 \tilde{e}^c - \\
 & - H_U^* m_{H_U}^2 H_U - H_D^* m_{H_D}^2 H_D - H_U \tilde{q}^* A_U \tilde{u}^c - \\
 & - H_D \tilde{q}^* A_U \tilde{d}^c - H_D \tilde{l}^* A_l \tilde{e}^c - B \mu H_U H_D .
 \end{aligned} \tag{36.37}$$

The last term is a quadratic term for the Higgs scalars. The three terms before it are the new trilinear scalar interactions, or  $A$ -terms. When the Higgses get VEVs, these terms also will induce sfermion masses (mixing left and right scalars). Finally,  $m_q^2$  etc are  $3 \times 3$  matrices in generation space. So are the  $A$ -terms ( $A_U$  etc).

The values of the SUSY-breaking parameters are determined by the symmetry breaking theory and mainly by the mediation. You sometimes hear people criticize MSSM for having a hundred or so new parameters (the parameters of  $\mathcal{L}_{soft}$ ). These are all determined however by the symmetry breaking and the mediation scheme. Often these involve very few new parameters (only one in anomaly mediation and two in minimal gauge mediation).

Note that the parameters of  $\mathcal{L}_{soft}$ , where all the interesting physics lies, are the only freedom we have. These parameters determine the spectrum of squarks, sleptons, gauginos, and therefore the way SUSY manifests itself.

The gaugino masses and  $A$ -terms break the  $U(1)_R$  symmetry of the MSSM Lagrangian. There is a discrete symmetry left however. This remanent symmetry is called  $R$ -parity. Under  $R$ -parity, the gauginos, sfermions, and Higgsinos are odd, and all SM fields are even. Thus, when we supersymmetrize the SM without adding any new interactions, we have a new parity symmetry, which guarantees that the lightest superpartner is stable!

Before we go on, let us discuss one remaining problem. We have two massless Higgsinos in the theory. As we saw above, these do not get mass from SUSY breaking. So we must also include a SUSY mass term for them,

$$W = \mu H_U H_D. \quad (36.38)$$

## 36.4 Mediating the Breaking

There are many possibilities to mediate SUSY breaking. One is the *Gauge Mediated SUSY Breaking*. Another is gravity, which is the basis of *Anomaly Mediated SUSY Breaking*. Planck-suppressed interactions, which are also associated with gravity, are at the basis of *Gravity Mediated SUSY Breaking*. Even Yukawa-like interactions can do the job.

### 36.4.1 Gauge Mediated Breaking

Gauge interactions are the ones we know best. Therefore gauge mediation gives full, concrete, and often fully calculable SUSY extensions of the SM.

We can start with a toy example to illustrate how things work. We saw in

the O’Raifeartaigh model,

$$W = \phi(\phi_1^2 - f) + m\phi_1\phi_2 . \quad (36.39)$$

Recall that this model breaks SUSY. The spectrum of the model contains a supermultiplet with SUSY-breaking mass splittings: a fermion of mass  $m$ , and scalars of masses-squared,  $m^2 + 2f$  and  $m^2 - 2f$ .

Let us complicate the model slightly, by considering five fields,  $\phi$ ,  $\phi_{1\pm}$  and  $\phi_{2\pm}$ , with the superpotential,

$$W = \phi(\phi_{1+}\phi_{1-} - f) + m\phi_{1+}\phi_{2-} + m\phi_{1-}\phi_{2+} . \quad (36.40)$$

Now the model has a  $U(1)$  symmetry, under which  $\phi$  has charge zero, and  $\phi_{i\pm}$  ( $i = 1, 2$ ) has charge  $\pm 1$ . It is easy to see that SUSY is still broken. Again we have supermultiplets with SUSY breaking splittings between fermions and bosons. Now let us promote the  $U(1)$  symmetry to a gauge symmetry, and identify it with hypercharge.

Another way to think about this is the following. Add to the SM the fields  $\phi$ ,  $\phi_{1\pm}$  and  $\phi_{2\pm}$  of hypercharges 0,  $\pm 1$ , respectively, with the superpotential (36.40). Now consider a squark. It is charged under hypercharge, so it couples to these split supermultiplets. Therefore, a squark mass is generated!

Minimal Gauge Mediation Models are the simplest models of this type. Suppose we have a SUSY-breaking model with chiral supermultiplets  $\Phi_i$  and  $\bar{\Phi}_i$  ( $i = 1, 2, 3$ ), such that the fermions  $\psi_{\Phi_i}$  and  $\psi_{\bar{\Phi}_i}$  combine into a Dirac fermion of mass  $M$ , and the scalars have masses-squared  $M^2 \pm F$  (with  $F < M^2$ ). Now identify  $i$  as an  $SU(3)$  color index. Thus  $\Phi$  is a 3 of  $SU(3)$  and  $\bar{\Phi}$  is a  $\bar{3}$  of  $SU(3)$ . These fields have SUSY-breaking masses. The gluino talks to the  $\Phi$ ’s directly and therefore gets mass at one loop.

The squarks talk to the gluino and therefore get mass at two loops. We have a gluino mass,

$$m_{\tilde{g}} \sim \frac{\alpha}{4\pi} \frac{F}{M} + \mathcal{O}\left(\frac{F^2}{M^2}\right) \quad (36.41)$$

and a squark mass-squared at two loops:

$$m_{\tilde{q}} \sim \frac{\alpha^2}{(4\pi)^2} \frac{F^2}{M^2} + \mathcal{O}\left(\frac{F^4}{M^6}\right) , \quad (36.42)$$

where the coefficients (group theory factors) are omitted. We can infer these expressions for masses very simply since the masses arise at one or two loops there is the appropriate loop factor;

- the masses should vanish as  $F \rightarrow 0$ ;
- the masses should vanish as  $M \rightarrow \infty$ .

Gauge mediation is very elegant:

- the soft masses are determined by the gauge couplings;
- the squark matrices are flavor-blind ( $\propto 1_{3 \times 3}$  in flavor space);
- the gluino masses  $\sim$  squark masses;
- the only new parameters are  $F$  and  $M$ , and the overall scale is  $F/M$ .  
If want soft masses around the TeV, then  $F/M \sim 100$  TeV.

The new fields  $\Phi$  are the *messengers* of SUSY breaking. In order to give masses to all the MSSM fields we need messenger fields charged under  $SU(3)$ ,  $SU(2)$  and  $U(1)$ , e.g.  $N_5$  copies of  $(3, 1)_{-1/3} + (\bar{3}, 1)_{1/3}$  and  $(1, 2)_{-1/2} + (1, 2)_{1/2}$  (filling up a  $5 + \bar{5}$  of  $SU(5)$ ). This adds another parameter, namely the number of messengers,  $N_5$ .

The messenger scale  $M$  mainly enters through running. The soft masses are generated at the messenger scale. To calculate them at the TeV we need to include renormgroup effects. The gravitino mass in these models is

$$m_{3/2} = \frac{F_{eff}}{M_{Pl}}, \quad (36.43)$$

where  $F_{eff}$  is the dominant  $F$  term. Therefore,

$$m_{3/2} \geq \frac{F}{M_{Pl}} \sim 100 \text{ TeV} \frac{M}{M_{Pl}} \quad (36.44)$$

and for low messenger scales, the gravitino can be very light ( $\sim eV$ ).

Minimal gauge mediation is just a simple example, mediation can in principle have a very different structure. The only defining feature is that the soft masses are generated by the SM gauge interactions. Generically we have:

- colored superpartners (gluinos, squarks) are heavier than non-colored (EW gauginos, sleptons) by a factor

$$\frac{\alpha_3}{\alpha_2} \quad \text{or} \quad \frac{\alpha_3}{\alpha_1}; \quad (36.45)$$

- the bino is the lightest gaugino and gaugino masses scale as

$$\alpha_3 : \alpha_2 : \alpha_1 ; \quad (36.46)$$

- to leading order, the  $A$  terms vanish at  $M$ ;
- the gravitino is light.

### 36.4.2 Gravity Mediation

With gauge mediation, we had to do some real work: add new fields, make sure they get SUSY-breaking masses couple them to the MSSM. But SUSY breaking is one place where we can expect an automatic result. Imagine we have, in addition to the SM, some symmetry breaking fields, e.g. the O’Raifeartaigh model. Since SUSY is a space-time symmetry, the SM fields should know this automatically. We would expect soft terms to be generated, suppressed by  $M_{Pl}$ . This is known as "gravity mediation".

We will discuss first the purest form of gravity mediation: anomaly mediation, and then what is commonly referred to as gravity mediation.

### 36.4.3 Anomaly Mediation

We assume that SUSY is broken by some fields that have no coupling to the SM. These fields are called the "hidden sector". The gravitino gets mass  $m_{3/2}$ . Would the MSSM "know" about SUSY breaking? Yes, at the quantum level, it is not scale-invariant, all the couplings (gauge, Yukawa) run – they are scale dependent. Therefore they are sensitive to the SUSY-breaking gravitino mass, and all the soft terms are generated. The gaugino masses are given by,

$$m_{1/2} = b \frac{\alpha}{4\pi} m_{3/2} , \quad (36.47)$$

where  $\alpha$  is the appropriate gauge coupling and  $b$  is the beta-function coefficient. Thus for  $SU(3)$ ,  $SU(2)$  and  $U(1)$  the parameter  $b$  equals to 3, -1 and -33/5, respectively.

Sfermions get masses proportional to their anomalous dimensions:

$$m_0^2 \sim \frac{1}{16\pi^2} (y^4 - y^2 g^2 + b g^4) m_{3/2}^2 . \quad (36.48)$$

For the first and second generation sfermions, we can neglect the Yukawas and,

$$m_0^2 \sim \frac{g^4}{16\pi^2} b m_{3/2}^2, \quad (36.49)$$

$A$  terms are generated too, proportional to the beta functions of the appropriate Yukawa.

This is amazing, all the soft terms are determined by just the SM couplings with one new parameter, the gravitino mass. It seems too good to be true. Indeed, while  $SU(3)$  is asymptotically free and  $b_3 > 0$ ,  $SU(2)$  and  $U(1)$  are not,  $b_2, b_1 < 0$ . Therefore the sleptons are tachyonic. There are various solutions to this problem, but the gaugino masses are fairly robust,

$$m_{\tilde{w}} : m_{\tilde{b}} : m_{\tilde{g}} : m_{3/2} \sim 1 : 3.3 : 10 : 370. \quad (36.50)$$

In this scenario, the wino is the lightest superpartner. Note that the gravitino is roughly a loop factor heavier than the SM superpartners.

## 36.5 Planck Suppressed Operators

Let us return to our basic setup. We expect to have higher-dimension operators, suppressed by  $M_{Pl}$ , that couple the symmetry breaking fields (the "hidden sector") and the MSSM fields. SUSY breaking leads to non-zero  $F$  terms (or  $D$  terms) for the symmetry breaking fields, so the higher-dimension operators coupling the two sectors will generate SUSY-breaking terms in the MSSM, with sfermion mass from

$$\frac{|F|^2}{M_{Pl}^2} \tilde{f}^\dagger \tilde{f} \quad (36.51)$$

and gaugino masses from

$$\frac{|F|}{M_{Pl}} \lambda^T \varepsilon \lambda. \quad (36.52)$$

You can think of these as mediated by tree-level exchange of Planck-scale fields.

Unlike in the previous two schemes, here we don't know the order-one coefficients. Consider for example the doublet-squarks. Their mass terms are,

$$c_{ij} \frac{|F|^2}{M_{Pl}^2} \tilde{q}_i^\dagger \tilde{q}_j, \quad (36.53)$$



where  $c_{ij}$  are order-one coefficients. Thus,

$$(m_{\tilde{q}}^2)_{ij} = c_{ij} m_0^2 . \quad \left( m_0 \equiv \frac{|F|}{M_{Pl}} \right) \quad (36.54)$$

In some models one assumes

$$c_{ij} = \delta_{ij} . \quad (36.55)$$

It is not easy to justify this: the Yukawas are presumably generated at this high scale, so there are flavor-dependent couplings in the theory.

Including the running to low scales,

$$\frac{d}{dt} m_{1/2} \propto \frac{\alpha}{4\pi} m_{1/2} , \quad (36.56)$$

we find that starting from a common gaugino mass at the GUT scale, the gaugino masses scale as

$$\alpha_3 : \alpha_2 : \alpha_1 , \quad (36.57)$$

just as in gauge mediation. Again the bino is the lightest superpartner. The gravitino mass is of order the superpartner masses in this case.

These are a few possibilities for mediating SUSY breaking but by no means an exhaustive list.

## 36.6 EW SSB and the Higgs Mass

In the MSSM we have two Higgses,  $H_U$  and  $H_D$ , which can get VEVs,

$$\langle H_U \rangle = \begin{pmatrix} v_U \\ 0 \end{pmatrix} , \quad \langle H_D \rangle = \begin{pmatrix} 0 \\ v_D \end{pmatrix} . \quad (36.58)$$

Let us start in the SUSY limit (with no  $mu$  term). The  $D$  term must vanish, so the VEVs must be equal,

$$D = 0 \quad \rightarrow \quad v_U = v_D . \quad (36.59)$$

The two Higgs fields contain eight real scalars. Of these, three are eaten by  $W^\pm$  and  $Z$ .

Consider the heavy  $Z$  supermultiplet. It contains a heavy gauge boson which has 3 physical polarizations, and therefore 3 bosonic degrees of freedoms and a Dirac fermion (4 degrees of freedoms). Therefore, in order to have the same number of fermion and boson degrees of freedoms there must be one more real scalar. This scalar comes from the Higgs fields. The same holds for the  $W^\pm$ . Thus, 3 real scalars "join" the heavy  $W^\pm$  and  $Z$  supermultiplets. In the limit of unbroken SUSY which we are assuming now, all of these fields have masses  $M_W$  or  $M_Z$ .

Thus, of the 8 real scalars in  $H_U$  and  $H_D$ , 2 neutral fields remain. One is the SM physical Higgs,  $h$ . The other must be there because we have SUSY, and  $h$  must reside in a chiral supermultiplet. As we saw above, this multiplet contains a complex scalar field.

Note that so far there is no potential for  $h$ . This is not surprising. We have not added any Higgs superpotential, so the Higgs could only have a quartic form  $V_D$ . But along the  $D$ -flat direction, the physical Higgs is massless. Thus its mass must come from SUSY breaking!

Fortunately SUSY is broken – we have soft terms. The Higgs potential comes from the following sources:

- The  $m\mu$  term:  $W = \mu H_U H_D$ ,

$$\delta V = |\mu|^2 |H_U|^2 + |\mu|^2 |H_D|^2 ; \quad (36.60)$$

- The Higgs soft masses:

$$\delta V = \tilde{m}_{H_U}^2 |H_U|^2 + \tilde{m}_{H_D}^2 |H_D|^2 , \quad (36.61)$$

so we need  $m_{H_U}^2 < 0$  and/or  $m_{H_D}^2 < 0$ ;

- The  $B\mu$  term:

$$\delta V = B\mu H_U H_D + \text{hc} . \quad (36.62)$$

These are all quadratic terms;

- Then we have quartic terms:

$$\delta V = \frac{1}{2} g_2^2 D^I D^I + \frac{1}{2} g_1^2 D_Y D_Y , \quad (36.63)$$

where

$$D^I = H_U^\dagger \tau^I H_U - H_D^\dagger \tau^{I*} H_D , \quad (36.64)$$

and

$$D_Y = \sum_i Y_i \tilde{f}_i^\dagger \tilde{f}_i + \frac{1}{2}(H_U^\dagger H_U - H_D^\dagger H_D). \quad (36.65)$$

Recall we had two parameters, the two Higgs VEVs. We can trade them for:

1.  $\sqrt{v_U^2 + v_D^2}$ : determined by  $W$  mass to be 246 GeV;
2.  $\tan \beta \equiv v_U/v_D$ .

Requiring a minimum of the potential determines:

$$B\mu = \frac{1}{2}(m_{H_U}^2 + m_{H_D}^2 + 2\mu^2) \sin 2\beta, \quad (36.66)$$

$$\mu^2 = \frac{m_{H_D}^2 - m_{H_U}^2 \tan^2 \beta}{\tan^2 \beta - 1} - \frac{M_Z^2}{2}. \quad (36.67)$$

Thus, for given  $m_{H_U}^2$  and  $m_{H_D}^2$ :  $B\mu$  and  $\mu$  are determined, and we have two free parameters,  $\tan \beta$  and  $\text{sign}(\mu)$ .

Expanding around the VEVs we find that the various scalars from  $H_U$  and  $H_D$  have the following masses (squared),

$$\begin{aligned} H^\pm &: M_W^2 + M_A^2, & (\text{SUSY} : M_W^2) \\ H^0 &: \frac{M_Z^2 + M_A^2}{2} + \sqrt{\frac{(M_Z^2 + M_A^2)^2}{4} - m_A^2 M_Z^2 \cos^2 2\beta}, & (\text{SUSY} : M_Z^2) \\ A^0 &: M_A^2 = B\mu(\cot \beta + \tan \beta), & (\text{SUSY} : 0) \end{aligned} \quad (36.68)$$

and for the physical Higgs,

$$m_h^2 = \frac{1}{2}(M_Z^2 + M_A^2) - \frac{1}{2}\sqrt{(M_Z^2 + M_A^2)^2 - 4m_A^2 M_Z^2 \cos^2 2\beta}, \quad (36.69)$$

This is a *prediction*:

$$m_h \leq m_Z |\cos 2\beta| \leq M_Z. \quad (36.70)$$

So *the measurement of the Higgs mass provides the first quantitative test of the MSSM*. It fails. However, the result (36.69) is a tree-level result. There are large radiative corrections to this result, mainly from stop loops. In the decoupling limit

$$m_h^2 \sim m_Z^2 \cos^2 2\beta + \frac{3m_t^2}{4\pi^2 v^2} \left[ \log \frac{M_S^2}{m_t^2} + \frac{X_t^2}{M_S^2} \right], \quad (36.71)$$

where

$$\begin{aligned} X_t &= A_t - \mu \cot \beta, & (\text{the LR stop mixing}) \\ M_S &= \sqrt{\tilde{m}_{\tilde{t}_1} \tilde{m}_{\tilde{t}_2}}, & (\text{the average stop mass}) \end{aligned} \quad (36.72)$$

This can raise Higgs mass to around 130 – 150 GeV. Thus for a 126 GeV Higgs we need heavy stops and/or large stop  $A$  terms. This is not very attractive. We wanted SUSY to solve the fine-tuning problem, for which we need light stops. So the large Higgs mass typically implies some fine-tuning.

There is another important caveat. So far we did not add in any Higgs potential on top of what the MSSM "gave us". Let us compare this to the SM. In the SM we add (by hand) a quartic Higgs potential, with a quartic coupling  $\lambda$ , to get the Higgs mass. Here we did not have to,  $D$ -terms give a quartic potential. As a result, there is no new parameter,  $\lambda = g$ . We could add a quartic interaction *a la* SM. To do that we must add at least one new field, a SM singlet  $S$ , with

$$W = \lambda S H_U H_D \quad \rightarrow \quad V = \lambda^2 (|H_U|^2 |H_D|^2 + \dots), \quad (36.73)$$

This model is called the Next to MSSM.

We can pursue the comparison to the SM at a deeper level. In the SM, we put in EW SSB by hand. We had to put in a negative mass-squared for the Higgs. In the MSSM, EW SSB can have a dynamical origin. Recall that we needed  $\tilde{m}_{H_U}^2 < 0$  or  $\tilde{m}_{H_D}^2 < 0$ . This happens (almost) automatically in SUSY theories, since the renormgroup equations drive the Higgs mass-squared negative! The crucial contribution is due to the large Yukawa coupling of the Higgs to stops.

Now let us see why this happens. Suppose we start with  $\tilde{m}_{H_U}^2 > 0$  at the SUSY breaking scale. The running gives

$$\frac{d}{dt} m_{H_U}^2 \sim -\frac{g^2}{16\pi^2} m_{1/2}^2 + \frac{y_t^2}{16\pi^2} \tilde{m}_t^2. \quad (36.74)$$

The negative Yukawa contribution wins because:

1. The top Yukawa is large compared to the  $SU(2)$  and  $U(1)$  gauge couplings;
2. The stop is colored, so the Yukawa contribution is enhanced by a color factor (=3).

Note that there are many scalars in the MSSM, so you could worry about their masses-squared driven negative by the renorgroup equations. However, the Higgs is special: it is an  $SU(3)$  singlet, so there is no large positive contribution from the gluino. Furthermore, it has an order-one Yukawa to the colored stop. Thus only the Higgs develops a VEV.

Let us summarize our results so far. Putting aside the unpleasant 126 GeV Higgs mass (which can be accounted for), SUSY gives a very beautiful picture. The MSSM has only log divergences: the quadratic divergence in the Higgs mass-squared is cancelled by superpartners at  $\tilde{m}$ . The tuning is then  $\sim M_Z^2/\tilde{m}^2$  and the hierarchy between the EW SSB scale and the Planck/GUT scale is *stabilized*.

Furthermore: starting with  $\tilde{m}_{H_U}^2 > 0$  in the UV, the running (from stops) drives it negative, and EW symmetry is broken, with a scale proportional to  $\tilde{m}$ .

Finally, we remark that with a symmetry breaking sector that breaks SUSY dynamically, the SUSY breaking scale is exponentially suppressed.  $\tilde{m}$  can naturally be around the TeV. In this case, the correct hierarchy between the EW SSB scale and the Planck/GUT scale is not only stabilized, but actually *generated*.

Returning to the Higgs mass with  $m_h = 126$  GeV the MSSM is stretched: we often need heavy stops which implies some level of tuning. More practically, discovery becomes more of a challenge.

Now that we understand SUSY breaking and EW symmetry breaking let us turn to the superpartner spectrum.

### 36.6.1 Neutralinos and Charginos

We have 4 neutral 2-component spinors: two gauginos and two Higgsinos,

$$\tilde{b}, \quad \tilde{W}^0, \quad \tilde{H}_D^0, \quad \tilde{H}_U^0, \quad (36.75)$$

with the mass matrix

$$\begin{pmatrix} M_1 & 0 & -g_1 v_D/\sqrt{2} & g_1 v_U/\sqrt{2} \\ 0 & M_2 & g_2 v_D/\sqrt{2} & -g_2 v_U/\sqrt{2} \\ -g_1 v_D/\sqrt{2} & g_2 v_D/\sqrt{2} & 0 & \mu \\ g_1 v_U/\sqrt{2} & -g_2 v_U/\sqrt{2} & \mu & 0 \end{pmatrix}. \quad (36.76)$$

Diagonalizing this matrix we find 4 neutralino mass eigenstates:  $\tilde{\chi}^0$  ( $i = 1, \dots, 4$ ).

Similarly, there are two charginos mass eigenstates which are combinations of the charged Higgsino and wino,  $\tilde{\chi}_i^\pm$  ( $i = 1, 2$ ).

### 36.6.2 Sfermion Spectrum

Consider for example the up squarks. There are 6 complex scalars:  $\tilde{u}_{Li}$  and  $\tilde{u}_{Ra}$ , with  $i, a = 1, 2, 3$  labeling the three generations. The mass (squared) matrix is therefore a  $6 \times 6$  matrix:

$$\begin{pmatrix} m_{LL}^2 & m_{LR}^2 \\ m_{LR}^{2\dagger} & m_{RR}^2 \end{pmatrix}, \quad (36.77)$$

where each of the blocks is  $3 \times 3$ .

Consider  $m_{U,LL}^2$ . It gets contributions from:

1. the MSSM Yukawa (supersymmetric);
2. the SUSY breaking mass-squared;
3. the  $D$ -term (because  $D \sim v_U^2 - v_D^2 + \tilde{q}^\dagger T q + \dots$ ).

Thus,

$$m_{U,LL}^2 = m_u^\dagger m_u + \tilde{m}_q^2 + D_U 1_{3 \times 3}, \quad (36.78)$$

and similarly for  $m_{U,RR}^2$ .

$m_{LR}^2$  gets contributions from:

1. the  $A$  term (SUSY breaking);
2. the  $\mu$  term:

$$\left| \frac{\partial W}{\partial H_D} \right|^2 \rightarrow \frac{\partial W}{\partial H_D} = \mu H_U + y_U q u^c. \quad (36.79)$$

So

$$m_{U,LR}^2 = v_U (A_U^* - y_U \mu \cot \beta). \quad (36.80)$$

The remaining sfermions (down-squarks, sleptons, sneutrinos) have a similar structure.

### 36.6.3 Flavor Structure

Now let us consider the sfermion flavor structure, starting with the up squarks as before. Work in the quark mass basis (up, charm, top): the Lagrangian contains the following:

- gaugino– $u_{Li}$ – $\tilde{u}_{Lj}$  couplings. Here the gaugino can be either a gluino, a wino or a bino. In our original Lagrangian, these are proportional to  $\delta_{ij}$ . We therefore say that the Lagrangian is given in the interaction basis. Note that since we are in the fermion mass basis, this defines the left up squark, charm squark and top squark (stop). For example, the left stop is the state that couples to a gluino and the doublet-top quark;
- gaugino– $u_{Ra}$ – $\tilde{u}_{Rb}$  couplings. Again, in our original Lagrangian, these are proportional to  $\delta_{ab}$ ;
- ...;
- The up squark  $6 \times 6$  mass matrix. This can in principle have an arbitrary structure. In particular, the various  $3 \times 3$  blocks need not be diagonal.

Diagonalizing the squark mass matrix, we get 6 mass eigenstates,  $\tilde{u}_I$ , with  $I = 1, \dots, 6$ . However, the gaugino–quark–squark couplings are no longer diagonal. Writing these in terms of the up-squark mass eigenstates we have in general

$$K_{iI} \tilde{g} u_{Li} - \tilde{u}_I . \quad (36.81)$$

These mix the different generations, and  $K_{iI}$  are the quark-squark mixing parameters. Each squark (mass state) is a composition of the different flavor states.

Are sfermions degenerate? Is  $m_{U,LL}^2 \propto 1$ ? That depends on the mediation of SUSY. However, we do not understand the structure of fermion masses. In fact this structure is very strange, suggesting a fundamental theory of flavor. If there is such a theory, it will also control the structure of  $m_{U,LL}^2$  and the other sfermion mass-matrices.

### 36.6.4 $R$ -parity Violating Couplings

So far, we merely generalized the SM gauge and Yukawa interactions. In the SM the Yukawa couplings were the only renormalizable couplings allowed. Now however, there are new renormalizable Yukawa-like interactions,

$$W = \lambda_{ijk} L_i L_j e_k^c + \lambda'_{ijk} L_i Q_j d_k^c + \lambda''_{ijk} u_i^c d_j^c d_k^c . \quad (36.82)$$

These are the only terms we can add, nothing else is gauge invariant. These terms are problematic. The first two terms break lepton number, the third breaks baryon number. If they are all there we would get proton decay! Note also that all these new terms break  $R$ -parity. If we impose  $R$ -parity, these dangerous terms are forbidden, and proton decay can only arise from higher-dimension operators, much like in the SM. But this may be overly restrictive. Certain flavor patterns of  $R$ -parity breaking operators are viable.

**Exercise 36.1:** Find the components of a chiral superfield, one of the irreducible representations of simplest SUSY.

**Exercise 36.2:** Write most general Lagrangian for the simplest chiral superfield (Wess-Zumino model).

**Exercise 36.3:** Consider a chiral superfield of charge coupled to an Abelian vector superfield. Write down the  $D$ -term part of the scalar potential. Show that a non-vanishing vacuum expectation value of  $D$  (the auxiliary field of the vector superfield) can break SUSY.

**Exercise 36.4:** Construct the most general superpotential for MSSM.

**Exercise 36.5:** Derive the scalar potential for the Higgs fields in the MSSM. There are two contributions from the superpotential, three soft-breaking terms and two  $D$  terms.

**Exercise 36.6:** Show that performing  $SU(2)$  rotations allows to rotate the VEVs to the neutral components in the Higgs potential of the MSSM.

**Exercise 36.7:** Consider the structure of the Feynman rules of the MSSM. Recall what general Lorentz structures are possible: Gauge self-couplings; Fermion gauge interactions; Sfermion gauge interactions; All sorts of Yukawa couplings; Triple and quartic scalar couplings (never try to count).



## Part XIII

# Lecture – Grand Unification Theories



## Chapter 37

# Preliminaries

Not all of the symmetries of the SM, are actually seen in ordinary life, since some of them are "spontaneously broken". This means that while they are symmetries of the laws of physics, they are not symmetries of the vacuum. To see these symmetries we need to do experiments at very high energies, where the asymmetry of the vacuum has less effect. So, the behavior of particles at lower energies is like a shadow of the fundamental laws of physics, cast down from on high: a cryptic clue we must struggle to interpret.

It is reasonable to ask if this process continues. Could the symmetries of the SM be just a subset of all the symmetries in nature? Could they be the low energy shadows of laws still more symmetric?

A *Grand Unified Theory* (GUT) constitutes a guess at what these "more symmetric" laws might be. It is a theory with more symmetry than the SM, which reduces to the SM at lower energies. It is also, therefore, an attempt to describe the physics at higher energies.

GUTs are speculative physics. The SM has been tested in countless experiments. There is a lot of evidence that it is an incomplete theory, and some vague clues about what the next theory might be like, but so far there is no empirical evidence that any GUT is correct – and even some empirical evidence that some GUTs, like  $SU(5)$ , are incorrect.

Nonetheless, GUTs are interesting to theoretical physicists, because they allow us to explore some very definite ideas about how to extend the SM.

And because they are based almost entirely on the representation theory of compact Lie groups, the underlying physical ideas provide a marvelous playground for this beautiful area of mathematics. Amazingly, this beauty then becomes a part of the physics.

### 37.1 GUT Idea

The representation of the SM gauge group seems *ad hoc*. Why this one? Why all those seemingly arbitrary hypercharges floating around, mucking up some otherwise simple representations? Why do both leptons and quarks come in left- and right-handed varieties, which transform so differently? Why do quarks come in charges which are in units  $1/3$  times an electron's charge? Why are there the same number of quarks and leptons? GUTs can shed light on these questions, using only group representation theory.

GUT is a model in particle physics in which, at high energy, the three gauge interactions (or forces) of the SM which define the electromagnetic, weak, and strong interactions, are merged into one single force. This unified interaction is characterized by one larger gauge symmetry and thus several force carriers, but one unified coupling constant. Models that do not unify all interactions using one simple group as the gauge symmetry, but do so using semi-simple groups, can exhibit similar properties and are sometimes referred to as GUT as well. Unifying gravity with the other three interactions would provide a theory of everything, rather than a GUT. Nevertheless, GUTs are often seen as an intermediate step towards a theory of everything.

A GUT model consists of a gauge group which is a compact Lie group. A Yang-Mills action is given by an invariant symmetric bilinear form over its Lie algebra, a Higgs sector consists of a number of scalar fields taking on values within real/complex representations and chiral Weyl fermions take on values within a complex representation of this Lie group. The GUT group contains the SM group and the Higgs fields acquire VEVs leading to a SSB to the SM. The Weyl fermions represent matter.

If GUT is realized in nature, there is the possibility of a grand unification epoch in the early universe in which the fundamental forces are not yet distinct. The current wisdom is that we live in a broken phase in which the world looks  $SU(3)_C \times U(1)_Q$  invariant to us and the low-energy phenomena are governed by strong interactions and electrodynamics. Growing with the

energy we start to see the degrees of freedom of a new dynamics which can be interpreted as a renormalizable  $SU(2)_L \times U(1)_Y$  gauge theory spontaneously broken into  $U(1)_Q$ . Thus, in analogy to the  $U(1)_Q \rightarrow SU(2)_L \times U(1)_Y$  case, one can imagine that at higher energies the SM gauge group  $SU(3)_C \times SU(2)_L \times U(1)_Y$  is embedded in a simple group  $G$ .

The first implication of the GUT *ansatz* is that at some scale  $M_U \gg M_W$  the relevant symmetry is  $G$  and the  $g_3$ ,  $g_2$  and  $g_1$  coupling constants of  $SU(3)_C \times SU(2)_L \times U(1)_Y$  merge into a single gauge coupling  $g_U$ . The rather different values for  $g_3$ ,  $g_2$  and  $g_1$  at low-energy are then due to renormalization effects. Actually one of the most solid hints in favor of GUT is the fact that the running within the SM shows an approximate convergence of the gauge couplings around  $10^{15}$  GeV.

The GUT idea, though a bit speculative, may have a deep impact on the understanding of our low-energy world. Consider for instance some unexplained features of SM, like charge quantization or anomaly cancellation, which appear as the consequence of starting with an anomaly-free simple group. In the SM anomaly cancellation implies charge quantization, after taking into account the gauge invariance of the Yukawa couplings. This feature is lost when one adds a right-handed neutrino  $\nu_R$ , unless  $\nu_R$  is a Majorana particle.

The novel particles predicted by GUTs are expected to have masses around the GUT scale – just a few orders of magnitude below the Planck scale – and so will be well beyond the reach of any foreseen particle collider experiments. Therefore, the particles predicted by GUTs will be unable to be observed directly. However, the effects of GUT might be detected through indirect observations such as proton decay, electric dipole moments of elementary particles, or the properties of neutrinos.

GUT is reminiscent of the unification of electric and magnetic forces by Maxwell's theory of electromagnetism in XIX century, but its physical implications and mathematical structure are qualitatively different. GUTs which aim to be completely realistic are quite complicated, even compared to the SM, because they need to introduce additional fields and interactions, or even additional dimensions of space. The main reason for this complexity lies in the difficulty of reproducing the observed fermion masses and mixing angles which may be related to an existence of some additional family symmetries beyond the conventional GUTs. Due to this difficulty, and due to the lack of any observed effect so far, there is no generally accepted GUT.

Let us mention some model independent predictions of GUTs.

- *Charge Quantization.* The fact that the electric charges of electrons and protons seem to cancel each other exactly is essential for the existence of the macroscopic world as we know it, but this important property of elementary particles is not explained in the SM. While the description of strong and weak interactions within the SM is based on gauge symmetries governed by the simple symmetry groups  $SU(3)$  and  $SU(2)$  which allow only discrete charges, the remaining component, the weak hypercharge interaction is described by an abelian symmetry  $U(1)$  which in principle allows for arbitrary charge assignments.

The observed charge quantization, namely the fact that all known elementary particles carry electric charges which appear to be exact multiples of the "elementary" charge, has led to the idea that hypercharge interactions and possibly the strong and weak interactions might be embedded in one Grand Unified interaction described by a single, larger simple group containing the SM. This would automatically predict the quantized nature and values of all elementary particle charges. Since this also results in a prediction for the relative strengths of the fundamental interactions which we observe, in particular the weak mixing angle, Grand Unification ideally reduces the number of independent input parameters, but is also constrained by observations.

- *Baryon Asymmetry of the Universe.* Most importantly Grand Unification is not just a mere interpretation of our low-energy world, but it predicts new phenomena which are correlated with the existing ones. The most prominent of these is the instability of matter. The current lower bound on the proton lifetime is something like twenty three orders of magnitude bigger than the age of the Universe, namely  $\tau_p \gtrsim 10^{33\div 34}$  yr depending on the decay channel. This number is so huge that people started to consider baryon number as an exact symmetry of Nature. Nowadays we interpret it as an accidental global symmetry of the SM.

In the SM the baryonic current is anomalous and baryon number violation can arise from instanton transitions between degenerate  $SU(2)_L$  vacua which lead to  $\Delta B = \Delta L = 3$  interactions for three flavor families. The rate is estimated to be proportional to  $e^{-2\pi/\alpha_2} \sim e^{-173}$  and thus phenomenologically irrelevant. This also means that as soon as we extend the SM there is the chance to introduce baryon violating interactions. Gravity itself could be responsible for the breaking of baryon

number. However, among all the possible frameworks there is only one theory which predicts a proton lifetime close to its experimental limit and this is GUT. Indeed we can roughly estimate it by dimensional arguments. The exchange of a baryon-number-violating vector boson of mass  $M_U$  yields something like

$$\tau_p \sim \frac{1}{\alpha_U} \frac{M_U^4}{m_p^5}, \quad (37.1)$$

and one discovers that  $\tau_p \gtrsim 10^{33}$  yr corresponds to  $M_U \gtrsim 10^{15}$  GeV. Notice that the gauge running is sensitive to the log of the scale. This means that a 10% variation on the gauge couplings at the EW scale induces a 100% one on  $M_U$ . Were the apparent unification of gauge couplings in the window  $10^{15 \div 18}$  GeV just an accident, then Nature would have played a bad trick on us.

- *Magnetic Monopoles.* Another firm prediction of GUTs are magnetic monopoles. Each time a simple gauge group  $G$  is broken to a subgroup with a  $U(1)$  factor there are topologically nontrivial configurations of the Higgs field which leads to stable monopole solutions of the gauge potential. For instance  $SU(5)$  breaking generates monopoles with magnetic charge  $Q_m = 2\pi/e$  and mass  $M_m = M_U/\alpha_U$ . The central core of a GUT monopole contains the fields of the superheavy gauge bosons which mediate proton decay, so one expects that baryon number can be violated in baryon-monopole scattering. Quite surprisingly it was found that these processes are not suppressed by powers of the unification mass, but have a cross section typical of the strong interactions.

Though GUT monopoles are too massive to be produced at accelerators, they could have been produced in the early universe as topological defects arising via the Kibble mechanism during a symmetry breaking phase transition. Experimentally one tries to measure their interactions as they pass through matter. The strongest bounds on the flux of monopoles come from their interactions with the galactic magnetic field ( $\Phi < 10^{-16}$  cm<sup>-2</sup> sr<sup>-1</sup> sec<sup>-1</sup>) and the catalysis of proton decay in compact astrophysical objects ( $\Phi < 10^{-18 \div 29}$  cm<sup>-2</sup> sr<sup>-1</sup> sec<sup>-1</sup>).

- *Neutrino Masses.* The issue of neutrino masses caught the attention of particle physicists since a long time ago. The model independent way to parameterize them is to consider the SM as an effective field theory by writing all the possible operators compatible with gauge invariance.

Remarkably at the  $d = 5$  level there is only one operator

$$\frac{Y_\nu}{\Lambda_L} (\ell^T \sigma_2 H) C (H^T \sigma_2 \ell) , \quad (37.2)$$

where flavor indices are suppressed and  $C$  is the Dirac charge conjugation matrix. After EW symmetry breaking  $\langle H \rangle = v$  and neutrinos pick up a Majorana mass term

$$M_\nu = Y_\nu \frac{v^2}{\Lambda_L} . \quad (37.3)$$

The lower bound on the highest neutrino eigenvalue inferred from

$$\sqrt{\Delta m_{\text{atm}}} \sim 0.05 \text{ eV} \quad (37.4)$$

tells us that the scale at which the lepton number is violated is

$$\Lambda_L \lesssim Y_\nu \mathcal{O}(10^{14 \div 15} \text{ GeV}) . \quad (37.5)$$

Actually there are only three renormalizable UV completion of the SM which can give rise to the operator (37.2). They go under the name of type-I, type-II and type-III seesaw and are respectively obtained by introducing a fermionic singlet  $(1, 1, 0)_F$ , a scalar triplet  $(1, 3, +1)_H$  and a fermionic triplet  $(1, 3, 0)_F$ . These vector-like fields, whose mass can be identified with  $\Lambda_L$ , couple at the renormalizable level with  $\ell$  and  $H$  so that the operator (37.2) is generated after integrating them out. Since their mass is not protected by the chiral symmetry it can be super-heavy and can explain the smallness of neutrino masses.

Notice that this is still an effective field theory language and we cannot tell at this level if neutrinos are light because  $Y_\nu$  is small or because  $\Lambda_L$  is large. It is clear that without a theory that fixes the structure of  $Y_\nu$  we don't have much to say about  $\Lambda_L$ . The other possibility is that we may probe experimentally the new degrees of freedom at the scale  $\Lambda_L$  in such a way to reconstruct the theory of neutrino masses. This could be the case for left-right symmetric theories where  $\Lambda_L$  is the scale of the  $V + A$  interactions.

As an example of a predictive theory which can fix both  $Y_\nu$  and  $\Lambda_L$  we can mention  $SO(10)$  unification. The most prominent feature of  $SO(10)$  is that a SM fermion family plus a right-handed neutrino fit into a single 16-dimensional spinorial representation. This readily implies that  $Y_\nu$  is correlated to the charged fermion Yukawas. At the



same time  $\Lambda_L$  can be identified with the  $B - L$  generator of  $SO(10)$ , and its breaking scale,  $M_{B-L} \lesssim M_U$ , is subject to the constraints of gauge coupling unification.

To summarize, the model independent predictions of GUTs are proton decay, magnetic monopoles and charge quantization (and their deep connection). However once we have a specific model we can do even more. For instance the huge ratio between the unification and the EW scale,  $M_U/M_W \sim 10^{13}$ , reminds us about the well-established hierarchy among the masses of charged fermions and those of neutrinos,  $m_f/m_\nu \sim 10^{7 \div 13}$ . This analogy hints also to a possible connection between GUTs and neutrino masses.

### 37.1.1 Historical Remarks

Historically, the first true GUT which was based on the simple Lie group  $SU(5)$ , was proposed by Georgi and Glashow in 1974. The Georgi-Glashow model was preceded by the semi-simple Lie algebra by Salam and Pati, who pioneered the idea to unify gauge interactions. The acronym GUT was first coined in 1978 by CERN researchers Ellis, Buras, Gaillard and Nanopoulos, however in the final version of their paper they opted for the less anatomical GUM (Grand Unification Mass). Nanopoulos later that year was the first to use the acronym in a paper.

We can mention the following historical phases:

- 1974–1986: Golden age of GUTs. These are the years of the foundation in which the fundamental aspects of the theory are worked out. The first estimate of the proton lifetime in 1974 by Georgi, Quinn and Weinberg yields  $\tau_p \sim 10^{31}$  yr, amazingly close to the experimental bound of Reines and Crouch  $\tau_p \gtrsim 10^{30}$  yr. Hence, it was the great hope that proton decay is behind the corner.
- 1987 ÷ 1990: Great depression. The proton lifetime is pushed to  $\tau_p \gtrsim 10^{27}$  yr. Neither proton decay nor magnetic monopoles are observed so far. Emblematically the last workshop on pure GUTs is held in 1989.
- $\gtrsim 1991$ : SUSY-GUTs. The new data of the LEP collider at CERN seem to favor low-energy SUSY as a candidate for gauge coupling unification. From now on almost all the attention is caught by SUSY.

- $\gtrsim$  1998: Neutrino revolution. Starting from 1998 experiments begin to show that atmospheric and solar neutrinos change flavor. GUTs come back with a rationale for the origin of the sub-eV neutrino mass scale.
- $\gtrsim$  2010: LHC era. Has SUSY something to do with the EW scale? The lack of evidence for SUSY at the LHC would undermine SUSY-GUT scenarios. Back to pure GUTs?
- $\gtrsim$  2019: Next generation of proton decay experiments sensitive to  $\tau_p \sim 10^{34\div 35}$  yr. The future of GUTs relies heavily on that.

Despite the huge amount of work done so far, the situation does not seem very clear at the moment. Especially from a theoretical point of view no model of grand unification emerged as 'the' theory. The reason can be clearly attributed to the lack of experimental evidence on proton decay.

In such a situation a good guiding principle in order to discriminate among models and eventually falsify them is given by minimality, where minimality deals interchangeably with simplicity, tractability and predictivity. It goes without saying that minimality could have nothing to do with our world, but it is anyway the best we can do at the moment. It is enough to say that if one wants to have under control all the aspects of the theory the degree of complexity of some minimal GUT is already at the edge of the tractability.

Quite surprisingly after many years from first works, there is still no consensus on which is the minimal theory. Maybe the reason is also that minimality is not a universal and uniquely defined concept, admitting a number of interpretations. For instance it can be understood as a mere simplicity related to the minimum rank of the gauge group. This was indeed the remarkable observation of Georgi and Glashow:  $SU(5)$  is the unique rank-4 simple group which contains the SM and has complex representations. However nowadays we can say for sure that the Georgi-Glashow model in its original formulation is ruled out because it does not unify and in this model neutrinos are massive. Moved by this double issue of the Georgi-Glashow model, some minimal extensions which can cure at the same time both unification and neutrino masses have been proposed.

There is currently no hard evidence that nature is described by a GUT. The discovery of neutrino oscillations indicates that the SM is incomplete and has led to renewed interest toward certain GUT. One of the few possible experimental tests of certain GUT is proton decay and also fermion masses.

There are a few more special tests for SUSY GUT. However, minimum proton lifetimes from research (at or exceeding the  $10^{34} - 10^{35}$  year range) have ruled out simpler GUTs and most non-SUSY models. The maximum upper limit on proton lifetime (if unstable), is calculated at  $6 \times 10^{39}$  years for SUSY models and  $1.4 \times 10^{36}$  years for minimal non-SUSY GUTs.

The gauge coupling strengths of QCD, the weak interaction and hypercharge seem to meet at a common length scale called the GUT scale and equal approximately to  $10^{16}$  GeV, which is slightly suggestive. This interesting numerical observation is called the gauge coupling unification, and it works particularly well if one assumes the existence of superpartners of the SM particles. Still it is possible to achieve the same by postulating, for instance, that ordinary (non-SUSY)  $SO(10)$  models break with an intermediate gauge scale, such as the one of Pati-Salam group.

$SU(5)$  can be considered the prototype GUT where to study all the fundamental aspects of grand unification from proton decay to neutrino masses. From a more pragmatic point of view one could instead use predictivity as a measure of minimality. This singles out  $SO(10)$  as the best candidate. At variance with  $SU(5)$ , the fact that all the SM fermions of one family fit into the same representation makes the Yukawa sector of  $SO(10)$  much more constrained.

Notice that here we do not have in mind flavor symmetries, indeed the GUT symmetry itself already constrains the flavor structure just because some particles live together in the same multiplet. Certainly one could improve the predictivity by adding additional ingredients like local, global, continuous, or discrete symmetries on top of the GUT symmetry. However, though there is nothing wrong with that, we feel that it would be a no-ending process based on assumptions which are difficult to disentangle from the unification idea. That is why we prefer to stick as much as possible to the gauge principle without further ingredients.

Establishing the minimal Higgs content needed for the GUT breaking is a basic question which has been addressed since the early days of the GUT program. Remarkably the general patterns of symmetry breaking in gauge theories with orthogonal and unitary groups were already analyzed in 1973/1974 by Li, contemporarily with the work of Georgi and Glashow.

Let us stress that the quest for the simplest Higgs sector is driven not only by aesthetic criteria but it is also a phenomenologically relevant issue related to

the tractability and the predictivity of the models. Indeed, the details of the symmetry breaking pattern, sometimes overlooked in the phenomenological analysis, give further constraints on the low-energy observables such as the proton decay and the effective SM flavor structure. For instance in order to assess quantitatively the constraints imposed by gauge coupling unification on the mass of the lepto-quarks responsible for proton decay it is crucial to have the scalar spectrum under control. Even in that case some degree of arbitrariness can still persist due to the fact that the spectrum can never be fixed completely but lives on a manifold defined by the vacuum conditions. This also means that if we aim to a falsifiable (predictive) GUT scenario, better we start by considering a minimal Higgs sector.

### 37.1.2 Gauge Coupling Unification

Up to the energy scale of  $10^2$  GeV, we are confident that the fundamental gauge symmetry of particle physics is that of the SM, i.e.  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . New physics may appear just above this scale, but there may also be a much higher energy scale where the three gauge groups of the SM become unified into some larger symmetry. This is the notion of grand unification and depends crucially on the values of the three observed gauge couplings at the EW scale, as well as the particle content of the assumed theory from that scale to the unification scale.

The basic tool for exploring the possibility of grand unification is the renormalization group evolution of the gauge couplings as a function of energy scale, given in one loop by

$$\frac{1}{\alpha_i(M_Z)} = \frac{1}{\alpha_i(M_U)} + \frac{b_i}{2\pi} \ln \frac{M_U}{M_Z} \approx 128, \quad (37.6)$$

where

$$\frac{1}{\alpha_1} = \frac{3 \cos^2 \theta_W}{5\alpha}, \quad \frac{1}{\alpha_2} = \frac{\sin^2 \theta_W}{\alpha}, \quad \frac{1}{\alpha_3(M_Z)} \approx \frac{25}{3},$$

$$\sin^2 \theta_W(M_Z) \approx 0.23, \quad \sin^2 \theta_W(M_U) = \frac{3}{8}, \quad \frac{1}{\alpha(M_Z)} \approx 128. \quad (37.7)$$

The coefficients  $b_i$  are obtained from the assumed particle content of the theory between  $M_Z$  and  $M_U$ .

The renorm-group running of the three gauge couplings in the SM has been found to nearly, but not quite, meet at the same point if the hypercharge is

normalized so that it is consistent with  $SU(5)$  or  $SO(10)$  GUTs, which are precisely the GUT groups which lead to a simple fermion unification. This is a significant result, as other Lie groups lead to different normalizations. However, if the SUSY extension (MSSM) is used instead of the SM, the match becomes much more accurate. In this case, the coupling constants of the strong and EW interactions meet at the grand unification energy, also known as the GUT scale:

$$\Lambda_{\text{GUT}} \approx 10^{16} \text{ GeV} . \quad (37.8)$$

It is commonly believed that this matching is unlikely to be a coincidence, and is often quoted as one of the main motivations to further investigate SUSY theories despite the fact that no SUSY partner particles have been experimentally observed. Also, most model builders simply assume SUSY because it solves the hierarchy problem – i.e. it stabilizes the EW Higgs mass against radiative corrections.

Note that the MSSM can allow the unification of gauge couplings but there remains a possible discrepancy, depending on the choice of inputs at the EW scale. In fact, this small discrepancy is taken seriously by proponents of specific GUTs and has been the subject of debates.

## 37.2 The Pati-Salam Model

At the end of this preliminary chapter let us consider a GUT model that is not so "grand" – its gauge group is not a simple Lie group, as it is for the  $SU(5)$  and  $SO(10)$  theories, considered in the following chapters.

In the Pati-Salam model (proposed in 1974) the unification is based on four quark color charges (dubbed red, green, blue and violet), instead of the conventional three, with the new "violet" quark being identified with the leptons.

The model also has Left-right symmetry and predicts the existence of a high energy right handed weak interaction with heavy  $W'$  and  $Z'$  bosons. Originally the fourth color was labelled "lilac" to alliterate with "lepton". The gauge group of the model is either  $SU(4) \times SU(2)_L \times SU(2)_R$ , or  $(SU(4) \times SU(2)_L \times SU(2)_R)/Z_2$  and the fermions form three families, each consisting of the representations  $(4, 2, 1)$  and  $(4, 1, 2)$ .

### 37.2.1 Left-Right Asymmetry in SM

In the SM there is an intrinsic lack of left-right symmetry, since only neutrinos belonging to  $SU(2)_L$  doublets are introduced, without any explanation of the phenomenological facts that neutrino masses are very small and the weak interactions are predominantly  $V - A$ . The situation can be schematically depicted in the following way

$$q = \begin{pmatrix} u_1 & u_2 & u_3 \\ d_1 & d_2 & d_3 \end{pmatrix}, \quad \ell = \begin{pmatrix} \nu \\ e \end{pmatrix}, \quad \begin{matrix} d^c = (d_1^c & d_2^c & d_3^c) \\ u^c = (u_1^c & u_2^c & u_3^c) \end{matrix}, \quad \begin{matrix} e^c \\ ? \end{matrix}, \quad (37.9)$$

where under  $SU(3)_C \times SU(2)_L \times U(1)_Y$ ,

$$\begin{aligned} q &= (3, 2, +\frac{1}{6}), & \ell &= (1, 2, -\frac{1}{2}), \\ d^c &= (\bar{3}, 1, +\frac{1}{3}), & u^c &= (\bar{3}, 1, -\frac{2}{3}), & e^c &= (1, 1, +1). \end{aligned} \quad (37.10)$$

Considering the SM as an effective theory, neutrino masses can be generated by a  $d = 5$  operator of the type

$$\frac{Y_\nu}{\Lambda_L} (\ell^T \sigma_2 H) C (H^T \sigma_2 \ell), \quad (37.11)$$

where  $C$  is the charge-conjugation matrix. After EW symmetry breaking,  $\langle H \rangle = v$ , neutrinos pick up a Majorana mass term  $M_\nu \nu^T C \nu$  with

$$M_\nu = Y_\nu \frac{v^2}{\Lambda_L}. \quad (37.12)$$

The lower bound on the highest neutrino eigenvalue inferred from

$$\sqrt{\Delta m_{\text{atm}}} \sim 0.05 \text{ eV} \quad (37.13)$$

tells us that the scale at which the lepton number is violated is

$$\Lambda_L \lesssim Y_\nu \mathcal{O}(10^{14 \div 15} \text{ GeV}). \quad (37.14)$$

Notice that without a theory which fixes the structure of  $Y_\nu$  we don't have much to say about the scale  $\Lambda_L$ .

Actually, by exploiting the Fierz identity

$$(\sigma_i)_{ab} (\sigma_i)_{cd} = 2\delta_{ad} \delta_{cb} - \delta_{ab} \delta_{cd}, \quad (37.15)$$

one finds that the operator in (37.11) can be equivalently written in three different ways

$$\begin{aligned}
 & (\ell^T \sigma_2 H) C (H^T \sigma_2 \ell) = \\
 & = \frac{1}{2} (\ell^T C \sigma_2 \sigma_i \ell) (H^T \sigma_2 \sigma_i H) = \\
 & = - (\ell^T \sigma_2 \sigma_i H) C (H^T \sigma_2 \sigma_i \ell) .
 \end{aligned} \tag{37.16}$$

Each operator in (37.16) hints to a different renormalizable UV completion of the SM. Indeed one can think those effective operators as the result of the integration of a heavy state with a renormalizable coupling of the type,

$$i (\ell^T \sigma_2 H) C \nu^c , \quad i (\ell^T C \sigma_2 \sigma_i \ell) \Delta_i , \quad i (\ell^T \sigma_2 \sigma_i H) C T_i , \tag{37.17}$$

where  $\nu^c$ ,  $\Delta_i$  and  $T_i$  are a fermionic singlet ( $Y = 0$ ), a scalar triplet ( $Y = +1$ ) and a fermionic triplet ( $Y = 0$ ). Notice that being  $\nu^c$ ,  $\Delta_i \oplus \Delta_i^*$  and  $T_i$  vector-like states their mass is not protected by the EW symmetry and it can be identified with the scale  $\Lambda_L$ , thus providing a rationale for the smallness of neutrino masses. This goes under the name of seesaw mechanism and the three options in (37.17) are classified respectively as type I, II and III seesaw.

### 37.2.2 Left-Right Symmetry

It is natural to introduce a SM-singlet fermion field  $\nu^c$ . In such a way the spectrum looks more "symmetric" and one can imagine that at higher energies the left-right symmetry is restored, in the sense that left and right chirality fermions are assumed to play an identical role prior to some kind of SSB.

The smallest gauge group that implement this idea is  $SU(3)_C \times SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times Z_2$ , where  $Z_2$  is a discrete symmetry which exchange  $SU(2)_L \leftrightarrow SU(2)_R$ . The field content of the theory can be schematically depicted as

$$\begin{aligned}
 q &= \begin{pmatrix} u_1 & u_2 & u_3 \\ d_1 & d_2 & d_3 \end{pmatrix} , & \ell &= \begin{pmatrix} \nu \\ e \end{pmatrix} , \\
 q^c &= \begin{pmatrix} d_1^c & d_2^c & d_3^c \\ -u_1^c & -u_2^c & -u_3^c \end{pmatrix} , & \ell^c &= \begin{pmatrix} e^c \\ -\nu^c \end{pmatrix} ,
 \end{aligned} \tag{37.18}$$

where under  $SU(3)_C \times SU(2)_L \times SU(2)_R \times U(1)_{B-L}$ ,

$$\begin{aligned} q &= (3, 2, 1, +\frac{1}{3}) , & \ell &= (1, 2, 1, -1) , \\ q^c &= (\bar{3}, 1, 2^*, -\frac{1}{3}) , & \ell^c &= (1, 1, 2^*, +1) . \end{aligned} \quad (37.19)$$

Given this embedding of the fermion fields one readily verifies that the electric charge formula takes the expression

$$Q = T_L^3 + T_R^3 + \frac{B-L}{2} . \quad (37.20)$$

Next we have to state the Higgs sector. In the early days of the development of left-right theories the breaking to the SM was minimally achieved by employing the following set of representations:

$$\delta_L = (1, 2, 1, +1) , \quad \delta_R = (1, 1, 2, +1) , \quad \Phi = (1, 2, 2^*, 0) . \quad (37.21)$$

However, in order to understand the smallness of neutrino masses it is better to consider

$$\Delta_L = (1, 3, 1, +2) , \quad \Delta_R = (1, 1, 3, +2) \quad (37.22)$$

in place of  $\delta_L$  and  $\delta_R$  above.

Choosing the matrix representation

$$\Delta_{L,R} = \frac{\sigma_i}{2} \Delta_{L,R}^i \quad (37.23)$$

for the  $SU(2)_{L,R}$  adjoint and defining the conjugate doublet

$$\tilde{\Phi} \equiv \sigma_2 \Phi^* \sigma_2 , \quad (37.24)$$

the transformation properties for the Higgs fields under  $SU(2)_L$  and  $SU(2)_R$  read

$$\begin{aligned} \Delta_L &\rightarrow U_L \Delta_L U_L^\dagger , & \Delta_R &\rightarrow U_R \Delta_R U_R^\dagger , \\ \Phi &\rightarrow U_L \Phi U_R^\dagger , & \tilde{\Phi} &\rightarrow U_L \tilde{\Phi} U_R^\dagger , \end{aligned} \quad (37.25)$$

and consequently we have

$$\begin{aligned} \delta_L \Delta_L &= [T_L^3, \Delta_L] , & \delta_L \Delta_R &= 0 , & \delta_L \Phi &= T_L^3 \Phi , & \delta_L \tilde{\Phi} &= T_L^3 \tilde{\Phi} , \\ \delta_R \Delta_L &= 0 , & \delta_R \Delta_R &= [T_R^3, \Delta_R] , & \delta_R \Phi &= -\Phi T_R^3 , & \delta_R \tilde{\Phi} &= -\tilde{\Phi} T_R^3 , \\ \delta_{B-L} \Delta_L &= 2\Delta_L , & \delta_{B-L} \Delta_R &= 2\Delta_R , & \delta_{B-L} \Phi &= 0 , & \delta_{B-L} \tilde{\Phi} &= 0 . \end{aligned} \quad (37.26)$$



Then, given the expression for the electric charge operator in (37.20), we can decompose these fields in the charge eigenstates,

$$\Delta_{L,R} = \begin{pmatrix} \Delta^+/\sqrt{2} & \Delta^{++} \\ \Delta^0 & -\Delta^+/\sqrt{2} \end{pmatrix}_{L,R},$$

$$\Phi = \begin{pmatrix} \phi_1^0 & \phi_1^+ \\ \phi_2^- & \phi_2^0 \end{pmatrix}, \quad \tilde{\Phi} = \begin{pmatrix} \phi_2^{0*} & -\phi_2^+ \\ -\phi_1^- & \phi_1^{0*} \end{pmatrix}. \quad (37.27)$$

In order to fix completely the theory one has to specify the action of the  $Z_2$  symmetry on the field content. There are two phenomenologically viable left-right discrete symmetries:  $Z_2^P$  and  $Z_2^C$ . They are defined as

$$Z_2^P : \begin{cases} \psi_L & \longleftrightarrow & \psi_R \\ \Delta_L & \longleftrightarrow & \Delta_R \\ \Phi & \longleftrightarrow & \Phi^\dagger \\ W_L^\mu & \longleftrightarrow & W_R^\mu \end{cases} \quad \text{and} \quad Z_2^C : \begin{cases} \psi_L & \longleftrightarrow & \psi_L^c \\ \Delta_L & \longleftrightarrow & \Delta_R^* \\ \Phi & \longleftrightarrow & \Phi^T \\ W_L^\mu & \longleftrightarrow & W_R^{\mu*} \end{cases}. \quad (37.28)$$

The implications of this two cases differ by the tiny amount of  $CP$  violation. Indeed when restricted to the fermion fields we can identify  $Z_2^P$  and  $Z_2^C$  respectively with  $P: \psi_L \rightarrow \psi_R$  and  $C: \psi_L \rightarrow \psi_L^c \equiv C\gamma_0\psi_R^*$ . In the former case the Yukawa matrices are hermitian while in the latter they are symmetric. So if  $CP$  is conserved (real couplings)  $Z_2^P$  and  $Z_2^C$  lead to the same predictions.

Notice that  $Z_2^C$  involves an exchange between spinors with the same chirality. In principle this would allow the embedding of  $Z_2^C$  into a gauge symmetry which commutes with the Lorentz group. The gauging is conceptually important since it protects the symmetry from unknown UV effects.

### 37.2.3 Symmetry Breaking

Let us consider now the symmetry breaking sector. From (37.27) we deduce that the SM-preserving vacuum directions are

$$\langle \Delta_{L,R} \rangle = \begin{pmatrix} 0 & 0 \\ v_{L,R} & 0 \end{pmatrix}, \quad \langle \Phi \rangle = \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}, \quad \langle \tilde{\Phi} \rangle = \begin{pmatrix} v_2^* & 0 \\ 0 & v_1^* \end{pmatrix}. \quad (37.29)$$

The minimization of the scalar potential shows that beside the expected left-right symmetric minimum  $v_L = v_R$ , we have also the asymmetric one

$$v_L \neq v_R, \quad v_L v_R = \gamma v_1^2, \quad (\text{in the approximation } v_2 = 0) \quad (37.30)$$

where  $\gamma$  is a combination of parameters of the Higgs potential. Since the discrete left-right symmetry is defined to transform  $\Delta_L \leftrightarrow \Delta_R$  ( $\Delta_L \leftrightarrow \Delta_R^*$ ) in the case of  $Z_2^P$  ( $Z_2^C$ ), the VEVs in (37.30) breaks it spontaneously. Phenomenologically we have to require  $v_R \gg v_1 \gg v_L$  which leads to the following breaking pattern

$$\begin{aligned} SU(3)_C \times SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times Z_2 &\xrightarrow{v_R} \\ &\xrightarrow{v_R} SU(3)_C \times SU(2)_L \times U(1)_Y \xrightarrow{v_1 \gg v_L} \\ &\xrightarrow{v_1 \gg v_L} SU(3)_C \times U(1)_Q, \end{aligned} \quad (37.31)$$

where the gauge hierarchy is set by the gauge boson masses  $M_{W_R}$ ,  $M_{Z_R} \gg M_{W_L}$  and  $M_{Z_L}$ . Let us verify this by computing  $M_{W_R}$  and  $M_{Z_R}$ . We start from the covariant derivative

$$D_\mu \Delta_R = \partial_\mu \Delta_R + i g_R [T_R^i, \Delta_R] (A_R^i)_\mu + i g_{B-L} \frac{B-L}{2} \Delta_R (A_{B-L})_\mu, \quad (37.32)$$

and the canonically normalized kinetic term

$$\text{Tr} (D_\mu \langle \Delta_R \rangle)^\dagger D^\mu \langle \Delta_R \rangle, \quad (37.33)$$

which leads to

$$M_{W_R}^2 = g_R v_R^2, \quad M_{Z_R}^2 = 2 (g_R^2 + g_{B-L}^2) v_R^2, \quad M_Y^2 = 0, \quad (37.34)$$

where

$$\begin{aligned} W_R^\pm &= \frac{A_R^1 \mp i A_R^2}{\sqrt{2}}, \\ Z_R &= \frac{g_R A_R^3 + g_{B-L} A_{B-L}}{\sqrt{g_R^2 + g_{B-L}^2}}, \\ Y &= \frac{g_{B-L} A_R^3 - g_R A_{B-L}}{\sqrt{g_R^2 + g_{B-L}^2}}. \end{aligned} \quad (37.35)$$

Given the relation  $g_Y^{-2} = g_R^{-2} + g_{B-L}^{-2}$  and the  $Z_2$  symmetry in (37.28) which implies  $g_R = g_L \equiv g$ , we obtain

$$M_{Z_R}^2 = \frac{2g^2}{g^2 - g_Y^2} M_{W_R}^2 \sim 2.6 M_{W_R}^2. \quad (37.36)$$

At the next stage of symmetry breaking ( $\langle \Phi \rangle \neq 0$  and  $\langle \Delta_L \rangle \neq 0$ ) an analogous calculation yields (in the approximation  $v_2 = 0$ ),

$$M_{W_L}^2 = \frac{1}{2}g^2(v_1^2 + 2v_L^2), \quad M_{Z_L}^2 = \frac{1}{2}(g^2 + g_Y^2)(v_1^2 + 4v_L^2), \quad M_A^2 = 0, \quad (37.37)$$

where

$$W_L^\pm = \frac{A_L^1 \mp iA_L^2}{\sqrt{2}}, \quad Z_L = \frac{g_L A_L^3 - g_Y A_Y}{\sqrt{g_L^2 + g_Y^2}}, \quad A = \frac{g_Y A_L^3 + g_L A_Y}{\sqrt{g_L^2 + g_Y^2}}. \quad (37.38)$$

Notice that in order to preserve  $\rho = 1$  at tree level, where

$$\rho \equiv \frac{M_{W_L}^2}{M_{Z_L}^2} \frac{g^2 + g_Y^2}{g^2}, \quad (37.39)$$

one has to require  $v_L \ll v_1$ .

On the other hand at energy scales between  $M_{W_L}$  and  $M_{W_R}$ ,  $SU(2)_L \times U(1)_Y$  is still preserved and (37.20) implies

$$\Delta T_R^3 = -\frac{1}{2}\Delta(B - L). \quad (37.40)$$

Since  $\Delta_R$  is an  $SU(2)_R$  triplet  $\Delta T_R^3 = 1$  and we get a violation of  $B - L$  by two units. Then two classes of  $B$  and  $L$  violating processes can arise:

- $\Delta B = 0$  and  $\Delta L = 2$  which imply Majorana neutrinos;
- $\Delta B = 2$  and  $\Delta L = 0$  which lead to neutron-antineutron oscillations.

### 37.2.4 The Origin of Neutrino Masses

The piece of lagrangian relevant for neutrinos is

$$\begin{aligned} \mathcal{L}_\nu = & iY_\Phi \ell^T C \sigma_2 \Phi \ell^c + i\tilde{Y}_\Phi \ell^T C \sigma_2 \tilde{\Phi} \ell^c + \\ & + iY_\Delta (\ell^T C \sigma_2 \Delta_L \ell + \ell^{cT} C \Delta_R^* \sigma_2 \ell^c) + \text{hc} . \end{aligned} \quad (37.41)$$

The invariance of (37.41) under the  $SU(2)_L \times SU(2)_R$  might not be obvious. So let us recall that, on top of the transformation properties in (37.25),  $\ell \rightarrow U_L \ell$ ,  $\ell^c \rightarrow U_R \ell^c$  and  $U_{L,R}^T \sigma_2 = \sigma_2 U_{L,R}^\dagger$ . After projecting (37.41) on

the SM vacuum directions and taking only the pieces relevant to neutrinos we get

$$\mathcal{L}_\nu = Y_\Phi \nu^T C \nu^c v_2 + \tilde{Y}_\Phi \nu^T C \nu^c v_1 + Y_\Delta (\nu^T C \nu v_L + \nu^{cT} C \nu^c v_R^*) + \text{hc} . \quad (37.42)$$

Let us take for simplicity  $v_2 = 0$  and consider real parameters. Then the neutrino mass matrix in the symmetric basis  $(\nu \ \nu^c)$  reads

$$\begin{pmatrix} Y_\Delta v_L & \tilde{Y}_\Phi v_1 \\ \tilde{Y}_\Phi^T v_1 & Y_\Delta v_R \end{pmatrix}, \quad (37.43)$$

and, given the hierarchy  $v_R \gg v_1 \gg v_L$ , the matrix in (37.43) is block-diagonalized by a similarity transformation involving the orthogonal matrix

$$\begin{pmatrix} 1 - \frac{1}{2}\rho\rho^T & \rho \\ -\rho^T & 1 - \frac{1}{2}\rho^T\rho \end{pmatrix}, \quad (37.44)$$

where  $\rho = \tilde{Y}_\Phi v_1 / Y_\Delta v_R$ . The diagonalization is valid up to  $\mathcal{O}(\rho^2)$  and yields

$$m_\nu = Y_\Delta v_L - \frac{\tilde{Y}_\Phi \tilde{Y}_\Phi^T v_1^2}{Y_\Delta v_R}. \quad (37.45)$$

The two contributions go under the name of type-II and type-I seesaw respectively. From the minimization of the potential, see (37.30), one gets  $v_L = \gamma v_1^2 / v_R$  and hence the effective neutrino mass matrix reads

$$m_\nu = \left( Y_\Delta \gamma - \frac{\tilde{Y}_\Phi \tilde{Y}_\Phi^T}{Y_\Delta} \right) \frac{v_1^2}{v_R}. \quad (37.46)$$

This equation is crucial since it shows a deep connection between the smallness of neutrino masses and the non-observation of  $V + A$  currents. Indeed in the limit  $v_R \rightarrow \infty$  we recover the  $V - A$  structure and  $m_\nu$  vanish.

Nowadays we know that neutrino are massive, but this information is not enough in order to fix the scale  $v_R$  because the detailed Yukawa structures are unknown. In this respect one can adopt two complementary approaches. From a pure phenomenological point of view one can hope that the  $V + A$  interactions are just behind the corner and experiments such as the LHC are probing right now the TeV region. Depending on the choice of the discrete left-right symmetry which can be either  $Z_2^P$  or  $Z_2^C$ , the strongest bounds on  $M_{W_R}$  are given by the  $K_L - K_S$  mass difference which yields  $M_{W_R} \gtrsim 4$  TeV in the case of  $Z_2^P$  and  $M_{W_R} \gtrsim 2.5$  TeV in the case of  $Z_2^C$ .

Alternatively one can imagine some well-motivated UV completion in which the Yukawa structure of the neutrino mass matrix is correlated to that of the charged fermions.

### 37.2.5 Lepton Number as a 4-th Color

One can go a little step further and imagine a partial unification scenario in which quarks and leptons belong to the same representations. The simplest implementation is obtained by collapsing the multiplets in (37.18) in the following way

$$Q = \begin{pmatrix} u_1 & u_2 & u_3 & \nu \\ d_1 & d_2 & d_3 & e \end{pmatrix}, \quad Q^c = \begin{pmatrix} d_1^c & d_2^c & d_3^c & e^c \\ -u_1^c & -u_2^c & -u_3^c & -\nu^c \end{pmatrix}, \quad (37.47)$$

so that  $SU(3)_C \times U(1)_{B-L} \subset SU(4)_C$  and the fermion multiplets transform as  $Q = (4, 2, 1)$  and  $Q^c = (\bar{4}, 1, 2^*)$  under  $SU(4)_C \times SU(2)_L \times SU(2)_R$ , which is known as the Pati-Salam group. Even in this case one can attach an extra discrete symmetry which exchange  $SU(2)_L \leftrightarrow SU(2)_R$ .

The Higgs sector of the model is essentially an extension of that of the left-right symmetric model. Indeed we have  $\Delta_L = (\bar{10}, 3, 1)$ ,  $\Delta_R = (\bar{10}, 1, 3)$  and  $\Phi = (1, 2, 2^*)$ . From the decomposition  $10 = 6(+2/3) \oplus 3(-2/3) \oplus 1(-2)$  under  $SU(4)_C \supset SU(3)_C \times U(1)_{B-L}$  and the expression for the electric charge operator in (37.20), we can readily see that  $\langle \Delta_R \rangle$  contains a SM-singlet direction and so the first stage of the breaking is given by

$$SU(4)_C \times SU(2)_L \times SU(2)_R \xrightarrow{\langle \Delta_R \rangle} SU(3)_C \times SU(2)_L \times U(1)_Y, \quad (37.48)$$

while the final breaking to  $SU(3)_C \times U(1)_Q$  is obtained by means of the bi-doublet VEV  $\langle \Phi \rangle$ . Analogously to the left-right symmetric case an EW triplet VEV  $\langle \Delta_L \rangle \ll \langle \Phi \rangle$  is induced by the Higgs potential and the conclusions about neutrino masses are the same.

### 37.2.6 Nuclear Stability

A peculiar feature of the Pati-Salam model is that the proton is stable in spite of the quark-lepton transitions due to the  $SU(4)_C$  interactions.

Let us consider first gauge interactions. The adjoint  $SU(4)_C$  decomposes as  $15 = 1(0) \oplus 3(+4/3) \oplus \bar{3}(-4/3) \oplus 8(0)$  under  $SU(3)_C \times U(1)_{B-L}$ . In particular the transitions between quark and leptons due to the Pati-Salam bosons,  $X_{PS} \equiv 3(+4/3)$  and  $\bar{X}_{PS} \equiv \bar{3}(-4/3)$ , come from the current interactions

$$\mathcal{L}_{PS} \supset \frac{g}{\sqrt{2}} \left( X_\mu^{PS} [\bar{u}\gamma^\mu \nu + \bar{d}\gamma^\mu e] + \bar{X}_\mu^{PS} [\bar{u}^c\gamma^\mu \nu^c + \bar{d}^c\gamma^\mu e^c] \right) + \text{hc}. \quad (37.49)$$

It turns out that the Lagrangian (37.49) has an accidental global symmetry  $G$ , where

$$\begin{aligned} G(X_{PS}) &= -\frac{2}{3}, & G(u) &= G(d) = +\frac{1}{3}, & G(\nu) &= G(e) = +1, \\ G(\bar{X}_{PS}) &= +\frac{2}{3}, & G(u^c) &= G(d^c) = -\frac{1}{3}, & G(\nu^c) &= G(e^c) = -1 \end{aligned} \quad (37.50)$$

$G$  is nothing but  $B + L$  when evaluated on the standard fermions. Thus, given that  $B - L$  is also a (gauge) symmetry, we conclude that both  $B$  and  $L$  are conserved by the gauge interactions.

The situation regarding the scalar interactions is more subtle. Actually in the minimal model there is a hidden discrete symmetry, which forbids all the  $\Delta B = 1$  transitions, like for instance  $qqql$ . A simple way to see it is that any operator of the type  $qqql \subset QQQQ$  and the  $Q^4$  term must be contracted with an  $\epsilon_{ijkl}$  tensor in order to form an  $SU(4)_C$  singlet. However, since the Higgs fields in the minimal model are either singlets or completely symmetric in the  $SU(4)_C$  space, they cannot mediate  $Q^4$  operators.

On the other hand  $\Delta B = 2$  transitions like neutron-antineutron oscillations are allowed and they proceed through  $d = 9$  operators of the type

$$\frac{\langle \Delta_R \rangle}{M_{\Delta_R}^6} (udd)(udd), \quad (37.51)$$

which are generated by the Pati-Salam breaking VEV,  $\langle \Delta_R \rangle$ . The fact that  $\langle \Delta_R \rangle$  can be pushed down relatively close to the TeV scale without making the proton to decay is phenomenologically interesting, since one can hope in testable neutron-antineutron oscillations. Present bounds on nuclear instability give  $\tau_N > 10^{32}$  yr, which translates into a bound on the neutron oscillation time  $\tau_{n-\bar{n}} > 10^8$  sec. Analogous limits come from direct reactor oscillations experiments. This sets a lower bound on the scale of  $\Delta B = 2$  non-SUSY ( $d = 9$ ) operators that varies from 10 to 300 TeV depending on model couplings. Thus neutron-antineutron oscillations probe scales far below the unification scale.

**Exercise 37.1:** What is GUT scale and how one can estimate its value?

**Exercise 37.2:** How the Pati-Salam model answers two questions about the SM: Why are quarks and leptons so similar? Why are left and right particles so different?

**Exercise 37.3:** Write down all possible dimension-6 operators which are invariant under the SM group and violate the baryon number conservation.

## Chapter 38

# Georgi-Glashow's $SU(5)$

In Georgi-Glashow model (proposed in 1974) the SM gauge groups  $SU(3) \times SU(2) \times U(1)$  are combined into a single simple gauge group –  $SU(5)$ . The unified group is then thought to be spontaneously broken into the SM subgroup below some high energy GUT scale.

Since the Georgi-Glashow model combines leptons and quarks into single irreducible representations, there exist interactions which do not conserve baryon number, although they still conserve  $B - L$  associated with the symmetry of the common representation. This yields a mechanism for proton decay, which rate can be predicted from the dynamics of the model.

Proton decay has not yet been observed experimentally, and the resulting lower limit on the lifetime of the proton contradicts the predictions of the model. However, the elegance of the Georgi-Glashow model has led particle physicists to use it as the foundation for more complex models which yield longer proton lifetimes, like  $SO(10)$  in basic, or SUSY variants.

### 38.1 Structure of the Model

If we require minimality (i.e. using of rank 4 groups, as in the SM) one reaches the remarkable conclusion that the only simple group with complex representations, which contains  $SU(3)_C \times SU(2)_L \times U(1)_Y$  as a subgroup, is  $SU(5)$ .

Let us consider the fundamental representation of  $SU(5)$  and denote it as a 5-dimensional vector  $5_i$  ( $i = 1, \dots, 5$ ). It is usual to embed  $SU(3)_C \times SU(2)_L$  in such a way that the first three components of  $5$  transform as a triplet of  $SU(3)_C$  and the last two components as a doublet of  $SU(2)_L$

$$5 = (3, 1) \oplus (1, 2) . \quad (38.1)$$

In the SM we have 15 Weyl fermions per family with quantum numbers

$$\begin{aligned} q &\sim (3, 2, +\frac{1}{6}) , & \ell &\sim (1, 2, -\frac{1}{2}) , \\ u^c &\sim (\bar{3}, 1, -\frac{2}{3}) , & d^c &\sim (\bar{3}, 1, +\frac{1}{3}) , & e^c &\sim (1, 1, +1) . \end{aligned} \quad (38.2)$$

How to embed these into  $SU(5)$ ? One would be tempted to try with a 15 of  $SU(5)$ . Actually from the anti-symmetric and symmetric decomposition of the product

$$5 \times 5 = 10_A \oplus 15_S , \quad (38.3)$$

and the fact that  $3 \times 3 = \bar{3}_A \oplus 6_S$  one concludes that some of the known quarks should belong to color sextets, which is not the case. So the next step is to try with  $5 \oplus 10$  or better with  $\bar{5} \oplus 10$  since there is no  $(3, 1)$  in the set of fields in (38.2). The decomposition of  $\bar{5}$  under  $SU(3)_C \times SU(2)_L \times U(1)_Y$  is simply

$$\bar{5} = (\bar{3}, 1, +\frac{1}{3}) \oplus (1, 2, -\frac{1}{2}) , \quad (38.4)$$

where we have exploited the fact that the hypercharge is a traceless generator of  $SU(5)$ , which implies the condition

$$3Y(d^c) + 2Y(\ell) = 0 . \quad (38.5)$$

So, up to a normalization factor, one may choose

$$Y(d^c) = \frac{1}{3} , \quad Y(\ell) = -\frac{1}{2} . \quad (38.6)$$

Then from (38.3) and (38.4) we get

$$10 = (5 \times 5)_A = (\bar{3}, 1, -\frac{2}{3}) \oplus (3, 2, +\frac{1}{6}) \oplus (1, 1, +1) . \quad (38.7)$$

Thus the embedding of a SM fermion family into  $\bar{5} \oplus 10$  reads

$$\bar{5} = \begin{pmatrix} d_1^c \\ d_2^c \\ d_3^c \\ e \\ -\nu \end{pmatrix} , \quad 10 = \begin{pmatrix} 0 & u_3^c & -u_2^c & u_1 & d_1 \\ -u_3^c & 0 & u_1^c & u_2 & d_2 \\ u_2^c & -u_1^c & 0 & u_3 & d_3 \\ -u_1 & -u_2 & -u_3 & 0 & e^c \\ -d_1 & -d_2 & -d_3 & -e^c & 0 \end{pmatrix} , \quad (38.8)$$



where we have expressed the  $SU(2)_L$  doublets as  $q = (u \ d)$  and  $\ell = (\nu \ e)$ . Notice in particular that the doublet embedded in  $\bar{5}$  is  $i\sigma_2\ell \sim \ell^*$ , the symbol ' $\sim$ ' stands for the fact that  $i\sigma_2\ell$  and  $\ell^*$  transform in the same way under  $SU(2)_L$ .

It may be useful to know how the  $SU(5)$  generators act on  $\bar{5}$  and 10. From the transformation properties

$$\bar{5}^i \rightarrow (U^\dagger)_k^i \bar{5}^k, \quad 10_{ij} \rightarrow U_i^k U_j^l 10_{kl}, \quad (38.9)$$

where  $U = e^{iT}$  and  $T^\dagger = T$ , we deduce that the action of the generators is

$$\delta \bar{5}^i = -T_k^i \bar{5}^k, \quad \delta 10_{ij} = \{T, 10\}_{ij}. \quad (38.10)$$

Already at this elementary level we can list a set of important features of  $SU(5)$  which are typical of any GUT.

### 38.1.1 Charge Quantization

The charges of quarks and leptons are related. Let us write the most general electric charge generator compatible with the  $SU(3)_C$  invariance and the  $SU(5)$  embedding

$$Q = \text{diag}(a, a, a, b, -3a - b), \quad (38.11)$$

i.e.  $\text{Tr}Q = 0$ . Then by applying (38.10) we find

$$\begin{aligned} Q(u) &= a + b, & Q(u^c) &= 2a, & Q(d) &= -2a - b, & Q(d^c) &= -a, \\ Q(e) &= -b, & Q(e^c) &= -3a, & Q(\nu) &= 3a + b, \end{aligned} \quad (38.12)$$

so that apart for a global normalization factor the charges do depend just on one parameter, which must be fixed by some extra assumption. Let us say we require  $Q(\nu) = 0$  that readily implies

$$Q(e^c) = -Q(e) = \frac{3}{2}Q(u) = -\frac{3}{2}Q(u^c) = -3Q(d) = 3Q(d^c) = b, \quad (38.13)$$

i.e. the electric charge of the SM fermions is a multiple of  $b$ .

### 38.1.2 Anomaly Cancellation

We know that in the SM all the gauge anomalies vanish. This property is preserved in  $SU(5)$  since  $\bar{5}$  and 10 have equal and opposite anomalies, so

that the theory is still anomaly free. In order to see this explicitly let us under the branching chain

$$SU(5) \supset SU(4) \times U(1)_A \supset SU(3) \times U(1)_A \times U(1)_B \quad (38.14)$$

decompose 5 and 10,

$$\begin{aligned} 5 &= 1(4) \oplus 4(-1) = 1(4, 0) \oplus 1(-1, 3) \oplus 3(-1, -1) , \\ 10 &= 4(3) \oplus 6(-2) = 1(3, 3) \oplus 3(3, -1) \oplus 3(-2, -2) \oplus \bar{3}(-2, -2) , \end{aligned} \quad (38.15)$$

where the  $U(1)$  charges are given up to a normalization factor. The anomaly  $\mathcal{A}(R)$  relative to a representation  $R$  is defined by

$$\text{Tr}\{T_R^a, T_R^b\}T_R^c = \mathcal{A}(R)d^{abc} , \quad (38.16)$$

where  $d^{abc}$  is a completely symmetric tensor. Then, given the properties

$$\mathcal{A}(R_1 \oplus R_2) = \mathcal{A}(R_1) + \mathcal{A}(R_2) \quad \text{and} \quad \mathcal{A}(\bar{R}) = -\mathcal{A}(R) , \quad (38.17)$$

it is enough to compute the anomaly of the  $SU(3)$  subalgebra of  $SU(5)$ ,

$$\begin{aligned} \mathcal{A}_{SU(3)}(\bar{5}) &= \mathcal{A}_{SU(3)}(\bar{3}) , \\ \mathcal{A}_{SU(3)}(10) &= \mathcal{A}_{SU(3)}(3) + \mathcal{A}_{SU(3)}(3) + \mathcal{A}_{SU(3)}(\bar{3}) , \end{aligned} \quad (38.18)$$

in order to conclude that

$$\mathcal{A}(\bar{5} \oplus 10) = 0 . \quad (38.19)$$

We see that in  $SU(5)$  anomaly cancellation and charge quantization are closely related. Actually it is not a chance that in the SM anomaly cancellation implies charge quantization, after taking into account the gauge invariance of the Yukawa couplings.

### 38.1.3 Gauge Coupling Running

At some grand unification mass scale  $M_U$  the relevant symmetry is  $SU(5)$  and the  $g_3$ ,  $g_2$  and  $g_1$  coupling constants of  $SU(3)_C \times SU(2)_L \times U(1)_Y$  merge into one single gauge coupling  $g_U$ . The rather different values for  $g_3$ ,  $g_2$  and  $g_1$  at low-energy are then due to renormalization effects.

Before considering the running of the gauge couplings we need to fix the relative normalization between  $g_2$  and  $g_1$ , which enter the weak interactions,  $g_2 T_3 + g_1 Y$ . We define

$$\zeta = \frac{\text{Tr} Y^2}{\text{Tr} T_3^2}, \quad (38.20)$$

so that

$$Y_1 \equiv \frac{Y}{\sqrt{\zeta}} \quad (38.21)$$

is normalized as  $T_3$ . In a unified theory based on a simple group, the coupling which unifies is then ( $g' Y_1 = g_1 Y$ )

$$g' \equiv \sqrt{\zeta} g_1. \quad (38.22)$$

Evaluating the normalization over a  $\bar{5}$  of  $SU(5)$  one finds

$$\zeta = \frac{3 \left(\frac{1}{3}\right)^2 + 2 \left(-\frac{1}{2}\right)^2}{\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2} = \frac{5}{3}, \quad (38.23)$$

and thus one obtains the tree level matching condition

$$g_U \equiv g_3(M_U) = g_2(M_U) = g'(M_U). \quad (38.24)$$

At energies  $\mu < M_U$  the running of the fine-structure constants ( $\alpha_i \equiv g_i^2/4\pi$ ) is given by

$$\frac{1}{\alpha_i} = \frac{1}{\alpha_i(0)} - \frac{a_i}{2\pi} \log \left( \frac{\mu}{\mu_0} \right) \quad (38.25)$$

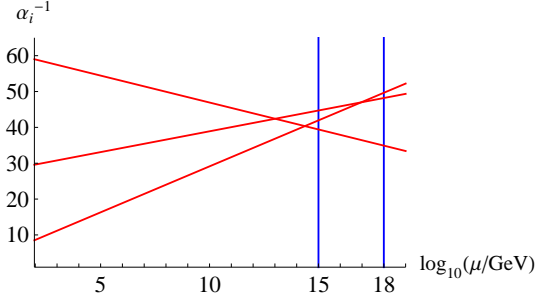
and the one-loop beta-coefficient for the SM reads

$$(a_3, a_2, a_1) = \left( -7, -\frac{19}{6}, \frac{41}{10} \right). \quad (38.26)$$

Starting from the experimental input values for the (consistently normalized) SM gauge couplings at the scale  $M_Z = 91.19$  GeV,

$$\alpha_1 \approx 0.017, \quad \alpha_2 \approx 0.034, \quad \alpha_3 \approx 0.118, \quad (38.27)$$

it is then a simple exercise to perform the one-loop evolution of the gauge couplings assuming just the SM as the low-energy effective theory. The result is depicted in the Figure below.



As we can see, the gauge couplings do not unify in the minimal framework, although a small perturbation may suffice to restore unification. In particular, threshold effects at the  $M_U$  scale (or below) may do the job, however depending on the details of the UV completion. By now this figure remains one of the most solid hints in favor of the grand unification idea. Indeed, being the gauge coupling evolution sensitive to the log of the scale, it is intriguing that they almost unify in a relatively narrow window,  $10^{15\div 18}$  GeV, which is still allowed by the experimental lower bound on the proton lifetime and a consistent effective QFT description without gravity.

## 38.2 Symmetry Breaking

The Higgs sector of the Georgi-Glashow model spans over the reducible  $5_H \oplus 24_H$  representation. These two fields are minimally needed in order to break the  $SU(5)$  gauge symmetry down to  $SU(3)_C \times SU(2)_L \times U(1)_Y$  and further to  $SU(3)_C \times U(1)_Q$ .

### 38.2.1 The First Stage

Let us concentrate on the first stage of the breaking which is controlled by the rank-conserving VEV  $\langle 24_H \rangle$ . The fact that the adjoint preserves the rank is easily seen by considering the action of the Cartan generators on the adjoint vacuum

$$\delta \langle 24_H \rangle 24_H^i_j = [T_{\text{Cartan}}, \langle 24_H \rangle^i_j]^i_j, \quad (38.28)$$

derived from the transformation properties of the adjoint

$$24^i_j \rightarrow (U^\dagger)^i_k U^l_j 24^k_l. \quad (38.29)$$

Since  $\langle 24_H \rangle$  can be diagonalized by an  $SU(5)$  transformation and the Cartan generators are diagonal by definition, one concludes that the adjoint preserves the Cartan subalgebra. The scalar potential is given by

$$V(24_H) = -m^2 \text{Tr} 24_H^2 + \lambda_1 (\text{Tr} 24_H^2)^2 + \lambda_2 \text{Tr} 24_H^4, \quad (38.30)$$

where for simplicity it is imposed the discrete symmetry  $24_H \rightarrow -24_H$ .

The minimization of the potential goes as follows:

First of all  $\langle 24_H \rangle$  is transformed into a real diagonal traceless matrix by means of an  $SU(5)$  transformation

$$\langle 24_H \rangle = \text{diag}(h_1, h_2, h_3, h_4, h_5), \quad (38.31)$$

where  $h_1 + h_2 + h_3 + h_4 + h_5 = 0$ . With  $24_H$  in the diagonal form, the scalar potential reads

$$V(24_H) = -m^2 \sum h_i^2 + \lambda_1 \left( \sum h_i^2 \right)^2 + \lambda_2 \sum h_i^4. \quad (38.32)$$

Since the  $h_i$ 's are not all independent, we need to use the Lagrangian multiplier  $\mu$  in order to account for the constraint  $\sum_i h_i = 0$ . The minimization of the potential  $V'(24_H) = V(24_H) - \mu \text{Tr} 24_H$  yields

$$\frac{\partial V'(24_H)}{\partial h_i} = -2m^2 h_i + 4\lambda_1 h_i \sum h_j^2 + 4\lambda_2 h_i^3 - \mu = 0. \quad (38.33)$$

Thus at the minimum all the  $h_i$ 's satisfy the same cubic equation

$$4\lambda_2 x^3 + (4\lambda_1 a - 2m^2) x - \mu = 0. \quad \left( a = \sum h_i^2 \right) \quad (38.34)$$

This means that the  $h_i$ 's can take at most three different values,  $\phi_1$ ,  $\phi_2$  and  $\phi_3$ , which are the three roots of the cubic equation. Note that the absence of the  $x^2$  term in the cubic equation implies that

$$\phi_1 + \phi_2 + \phi_3 = 0. \quad (38.35)$$

Let  $n_1$ ,  $n_2$  and  $n_3$  the number of times  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  appear in  $\langle 24_H \rangle$ ,

$$\langle 24_H \rangle = \text{diag}(\phi_1, \dots, \phi_2, \dots, \phi_3). \quad (n_1 \phi_1 + n_2 \phi_2 + n_3 \phi_3 = 0) \quad (38.36)$$

Thus  $\langle 24_H \rangle$  is invariant under  $SU(n_1) \times SU(n_2) \times SU(n_3)$  transformations. This implies that the most general form of symmetry breaking is  $SU(n) \rightarrow$

$SU(n_1) \times SU(n_2) \times SU(n_3)$  as well as possible  $U(1)$  factors (total rank is 4) which leave  $\langle 24_H \rangle$  invariant. To find the absolute minimum we have to use the relations

$$n_1\phi_1 + n_2\phi_2 + n_3\phi_3 = 0 \quad \text{and} \quad \phi_1 + \phi_2 + \phi_3 = 0 \quad (38.37)$$

to compare different choices of  $\{n_1, n_2, n_3\}$  in order to get the one with the smallest  $V(24_H)$ . It turns out that for the case of interest there are two possible patterns for the symmetry breaking:

$$SU(5) \rightarrow SU(3) \times SU(2) \times U(1), \quad \text{or} \quad SU(5) \rightarrow SU(4) \times U(1), \quad (38.38)$$

depending on the relative magnitudes of the parameters  $\lambda_1$  and  $\lambda_2$ . In particular for  $\lambda_1 > 0$  and  $\lambda_2 > 0$  the absolute minimum is given by the SM vacuum and the adjoint VEV reads

$$\langle 24_H \rangle = V \text{diag}(2, 2, 2, -3, -3). \quad (38.39)$$

Then the stability of the vacuum requires

$$\lambda_1 \left( \text{Tr} \langle 24_H \rangle^2 \right)^2 + \lambda_2 \text{Tr} \langle 24_H \rangle^4 > 0 \quad \Longrightarrow \quad \lambda_1 > -\frac{7}{30} \lambda_2 \quad (38.40)$$

and the minimum condition

$$\frac{\partial V(\langle 24_H \rangle)}{\partial V} = 0 \quad \Longrightarrow \quad 60V [-m^2 + 2V^2(30\lambda_1 + 7\lambda_2)] = 0 \quad (38.41)$$

yields

$$V^2 = \frac{m^2}{2(30\lambda_1 + 7\lambda_2)}. \quad (38.42)$$

Let us now write the covariant derivative

$$D_\mu 24_H = \partial_\mu 24_H + ig [A_\mu, 24_H], \quad (38.43)$$

where  $A_\mu$  and  $24_H$  are  $5 \times 5$  traceless hermitian matrices. Then from the canonical kinetic term,

$$\text{Tr} D_\mu \langle 24_H \rangle D^\mu \langle 24_H \rangle^\dagger = g^2 \text{Tr} [A_\mu, \langle 24_H \rangle] [\langle 24_H \rangle, A^\mu] \quad (38.44)$$

and the shape of the vacuum

$$\langle 24_H \rangle_j^i = h_j \delta_j^i, \quad (38.45)$$

where repeated indices are not summed, we can easily extract the gauge bosons mass matrix from the expression

$$g^2 [A_\mu, \langle 24_H \rangle]_j^i [\langle 24_H \rangle, A^\mu]_i^j = g^2 (A_\mu)_j^i (A^\mu)_i^j (h_i - h_j)^2 . \quad (38.46)$$

The gauge boson fields  $(A_\mu)_j^i$  having  $i = 1, 2, 3$  and  $j = 4, 5$  are massive,  $M_X^2 = 25g^2V^2$ , while  $i, j = 1, 2, 3$  and  $i, j = 4, 5$  are still massless.

Notice that the hypercharge generator commutes with the vacuum in (38.39) and hence the associated gauge boson is massless as well. The number of massive gauge bosons is then  $24 - (8 + 3 + 1) = 12$  and their quantum numbers correspond to the coset  $SU(5)/SU(3)_C \times SU(2)_L \times U(1)_Y$ . Their mass  $M_X$  is usually identified with the grand unification scale,  $M_U$ .

### 38.2.2 The Second Stage and Doublet-Triplet Splitting

The second breaking step,  $SU(3)_C \times SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \times U(1)_Q$ , is driven by a  $5_H$  where

$$5_H = \begin{pmatrix} T \\ H \end{pmatrix} , \quad (38.47)$$

decomposes into a color triplet  $T$  and an  $SU(2)_L$  doublet  $H$ . The latter plays the same role of the Higgs doublet of the SM.

The most general potential containing both  $24_H$  and  $5_H$  can be written as

$$V = V(24_H) + V(5_H) + V(24_H, 5_H) , \quad (38.48)$$

where  $V(24_H)$  is defined in (38.30),

$$V(5_H) = -\mu^2 5_H^\dagger 5_H + \lambda \left( 5_H^\dagger 5_H \right)^2 , \quad (38.49)$$

and

$$V(24_H, 5_H) = \alpha 5_H^\dagger 5_H \text{Tr} 24_H^2 + \beta 5_H^\dagger 24_H^2 5_H . \quad (38.50)$$

Again we have imposed for simplicity the discrete symmetry  $24_H \rightarrow -24_H$ .

It is instructive to compute the mass of the doublet  $H$  and the triplet  $T$  in the SM vacuum just after the first stage of the breaking

$$M_H^2 = -\mu^2 + (30\alpha + 9\beta)V^2 , \quad M_T^2 = -\mu^2 + (30\alpha + 4\beta)V^2 . \quad (38.51)$$

The gauge hierarchy  $M_X \gg M_W$  requires that the doublet  $H$ , containing the would-be Goldstone bosons eaten by the  $W$  and the  $Z$  and the physical Higgs boson, live at the  $M_W$  scale. This is unnatural and can be achieved at the prize of a fine-tuning of one part in  $\mathcal{O}(M_X^2/M_W^2) \sim 10^{26}$  in the expression for  $M_H^2$ . If we follow the principle that only the minimal fine-tuning needed for the gauge hierarchy is allowed then  $M_T$  is automatically kept heavy. This goes under the name of doublet-triplet splitting. Usually, but not always, a light triplet is very dangerous for the proton stability since it can couple to the SM fermions in such a way that baryon number is not anymore an accidental global symmetry of the low-energy Lagrangian.

A final comment about the radiative stability of the fine-tuning is in order. While SUSY helps in stabilizing the hierarchy between  $M_X$  and  $M_W$  against radiative corrections, it does not say much about the origin of this hierarchy. Other mechanisms have to be devised to render the hierarchy natural. In a non-SUSY scenario one needs to compute the mass of the doublet in (38.51) within a 13-loop accuracy in order to stabilize the hierarchy.

### 38.3 Proton Decay

The theory predicts that protons eventually decay. The most emblematic contribution to proton decay is due to the exchange of super-heavy gauge bosons which belong to the coset  $SU(5)/SU(3)_C \times SU(2)_L \times U(1)_Y$ .

Let us denote the matter representations of  $SU(5)$  as

$$\bar{5} = (\psi_\alpha, \psi_i) \ , \quad 10 = (\psi^{\alpha\beta}, \psi^{\alpha i}, \psi^{ij}) \ , \quad (38.52)$$

where the Greek and Latin indices run respectively from 1 to 3 ( $SU(3)_C$  space) and 1 to 2 ( $SU(2)_L$  space).

Analogously the adjoint 24 can be represented as

$$24 = (X_\beta^\alpha, X_j^i, X_\alpha^\alpha - \frac{3}{2}X_i^i, X_i^\alpha, X_\alpha^i) \ , \quad (38.53)$$

from which we can readily recognize the gauge bosons associated to the SM unbroken generators,  $(8, 1) \oplus (3, 1) \oplus (1, 1)$ , and the two super-heavy leptoquark gauge bosons,  $(3, 2) \oplus (\bar{3}, 2)$ .



Let us consider now the gauge action of  $X_i^\alpha$  on the matter fields,

$$\begin{aligned}\psi_\alpha &\rightarrow \psi_i, & (d^c \rightarrow \nu, e) \\ \psi^{\beta i} &\rightarrow \psi^{\beta\alpha}, & (d, u \rightarrow u^c) \\ \psi^{ij} &\rightarrow \psi^{\alpha j}. & (e^c \rightarrow u, d)\end{aligned}\tag{38.54}$$

Thus diagrams involving the exchange of a  $X_i^\alpha$  boson generate processes like

$$ud \rightarrow u^c e^c, \tag{38.55}$$

whose amplitude is proportional to the gauge boson propagator. After dressing the operator with a spectator quark  $u$ , we can have for instance the low-energy process  $p \rightarrow \pi^0 e^+$ , whose decay rate can be estimated by simple dimensional analysis

$$\Gamma(p \rightarrow \pi^0 e^+) \sim \frac{\alpha_U^2 m_p^5}{M_X^4}. \tag{38.56}$$

Using  $\tau(p \rightarrow \pi^0 e^+) > 8.2 \times 10^{33}$  years we extract the naive lower bound on the super-heavy gauge boson mass

$$M_X > 2.3 \times 10^{15} \text{ GeV}, \quad (\alpha_U^{-1} = 40) \tag{38.57}$$

which points directly to the GUT scale extrapolated by the gauge running (see e.g. the figure above).

Notice that  $B-L$  is conserved in the process  $p \rightarrow \pi^0 e^+$ . This selection rule is a general feature of the gauge induced proton decay and can be traced back to the presence of a global  $B-L$  accidental symmetry in the transitions of (38.54) after assigning  $B-L(X_i^\alpha) = 2/3$ .

## 38.4 Yukawa Sector and Neutrinos

The  $SU(5)$  Yukawa lagrangian can be written schematically as

$$\mathcal{L}_Y = \bar{5}_F Y_5 10_F 5_H^* + \frac{1}{8} \epsilon_5 10_F Y_{10} 10_F 5_H + \text{h.c.}, \tag{38.58}$$

where  $\epsilon_5$  is the 5-index Levi-Civita tensor. Here

$$\begin{aligned}\bar{5}_F Y_5 10_F 5_H^* &\equiv (\bar{5}_F)_m^{\alpha x} C_{xy} (Y_5)^{mn} (10_F)_{\alpha\beta n}^y (5_H^*)^\beta, \\ \epsilon_5 10_F Y_{10} 10_F 5_H &\equiv \epsilon^{\alpha\beta\gamma\delta\epsilon} (10_F)_{\alpha\beta m}^x C_{xy} (Y_{10})^{mn} (10_F)_{\gamma\delta n}^y (5_H)_\epsilon, \end{aligned}\tag{38.59}$$

where  $(\alpha, \beta, \gamma, \delta, \epsilon)$ ,  $(m, n)$  and  $(x, y)$  are respectively  $SU(5)$ , family and Lorentz indices.

After denoting the  $SU(5)$  representations synthetically as

$$\bar{5}_F = \begin{pmatrix} d^c \\ \epsilon_2 \ell \end{pmatrix}, \quad 10_F = \begin{pmatrix} \epsilon_3 u^c & q \\ -q^T & i\sigma_2 e^c \end{pmatrix}, \quad 5_H = \begin{pmatrix} T \\ H \end{pmatrix}, \quad (38.60)$$

where  $\epsilon_3$  is the 3-index Levi-Civita tensor, we project (38.58) over the SM components. This yields

$$\begin{aligned} \bar{5}_F Y_5 10_F 5_H^* &= i(d^c \ell \sigma_2^T) \begin{pmatrix} \epsilon_3 u^c & q \\ -q^T & i\sigma_2 e^c \end{pmatrix} \begin{pmatrix} T^* \\ H^* \end{pmatrix} \rightarrow \\ &\rightarrow d^c Y_5 q H^* + \ell Y_5 e^c H^*, \quad (38.61) \\ \frac{1}{8} \epsilon_5 10_F Y_{10} 10_F 5_H &\rightarrow \frac{1}{2} u^c (Y_{10} + Y_{10}^T) q H. \end{aligned}$$

After rearranging the order of the  $SU(2)_L$  doublet and singlet fields in the second line of (38.61), i.e.  $\ell Y_5 e^c H^* = e^c Y_5^T \ell H^*$ , one gets

$$Y_d = Y_e^T \quad \text{and} \quad Y_u = Y_u^T, \quad (38.62)$$

which shows a deep connection between flavor and the GUT symmetry (which is not related to a flavor symmetry).

The first relation in (38.62) predicts

$$m_b(M_U) = m_\tau(M_U), \quad m_s(M_U) = m_\mu(M_U), \quad m_d(M_U) = m_e(M_U) \quad (38.63)$$

at the GUT scale. So in order to test this relation one has to run the SM fermion masses starting from their low-energy values. While  $m_b(M_U) = m_\tau(M_U)$  is obtained in the MSSM with a typical 20 – 30% uncertainty, the other two relations are evidently wrong. By exploiting the fact that the ratio between  $m_d/m_e$  and  $m_s/m_\mu$  is essentially independent of renormalization effects, we get the scale free relation

$$\frac{m_d}{m_s} = \frac{m_e}{m_\mu}, \quad (38.64)$$

which is off by one order of magnitude.

Notice that  $m_d = m_e$  comes from the fact that the fundamental  $\langle 5_H \rangle$  breaks  $SU(5)$  down to  $SU(4)$  which remains an accidental symmetry of the Yukawa sector. So one expects that considering higher dimensional representations

makes it possible to further break the remnant  $SU(4)$ . This is indeed what happens by introducing a  $45_H$  which couples to the fermions in the following way

$$\bar{5}_F 10_F 45_H^* + 10_F 10_F 45_H + \text{hc} . \quad (38.65)$$

The first operator leads to  $Y_d = -3Y_e$ , so that if both  $5_H$  and  $45_H$  are present more freedom is available to fit all fermion masses. Alternatively one can built an effective coupling

$$\frac{1}{\Lambda} \bar{5}_F 10_F (\langle 24_H \rangle 5_H^*)_{\overline{45}} , \quad (38.66)$$

which mimics the behavior of the  $45_H$ . If we take the cut-off to be the planck scale  $M_{Pl}$ , this nicely keeps  $b - \tau$  unification while corrects the relations among the first two families. However in both cases we loose predictivity since we are just fitting  $M_d$  and  $M_e$  in the extended Yukawa structure.

### 38.4.1 Neutrino Masses

Finally what about neutrinos? It turns out that the Georgi-Glashow model has an accidental global  $U(1)_G$  symmetry with the charge assignment

$$G(\bar{5}_F) = -\frac{3}{5} , \quad G(10_F) = +\frac{1}{5} , \quad G(5_H) = +\frac{2}{5} . \quad (38.67)$$

The VEV  $\langle 5_H \rangle$  breaks this global symmetry but leaves invariant a linear combination of  $G$  and a Cartan generator of  $SU(5)$ . It easy to see that any linear combination of  $G + \frac{4}{5}Y$ ,  $Q$ , and any color generators is left invariant. The extra conserved charge  $G + \frac{4}{5}Y$  when acting on the fermion fields is just  $B - L$ . Thus neutrinos cannot acquire neither a Dirac (because of the field content) nor a Majorana (because of the global  $B - L$  symmetry) mass term and they remain exactly massless even at the quantum level.

Going at the non-renormalizable level we can break the accidental  $U(1)_G$  symmetry. For instance global charges are expected to be violated by gravity and the simplest effective operator one can think of is

$$\frac{1}{M_{Pl}} \bar{5}_F \bar{5}_F 5_H 5_H . \quad (38.68)$$

However its contribution to neutrino masses is too much suppressed ( $m_\nu \sim \mathcal{O}(M_W^2/M_{Pl}) \sim 10^{-5}$  eV). Thus we have to extend the field content of

the theory in order to generate phenomenologically viable neutrino masses. Actually, the possibilities are many.

Minimally one may add an  $SU(5)$  singlet fermion field  $1_F$ . Then, through its renormalizable coupling  $\bar{5}_F 1_F 5_H$ , one integrates  $1_F$  out and generates an operator similar to that in (38.68), but suppressed by the  $SU(5)$ -singlet mass term which can be taken well below  $M_{Pl}$ .

A slightly different approach could be breaking the accidental  $U(1)_G$  symmetry by adding additional scalar representations. Let us take for instance a  $10_H$  and consider then the new couplings

$$\mathcal{L}_{10} \supset f \bar{5}_F \bar{5}_F 10_H + M 10_H 10_H 5_H . \quad (38.69)$$

Since  $G(\bar{5}_F) = -3/5$  and  $G(5_H) = +2/5$  there is no way to assign a  $G$ -charge to  $10_H$  in order to preserve  $U(1)_G$ . Thus we expect that loops containing the  $B - L$  breaking sources  $f$  and  $M$  can generate neutrino masses.

So what is wrong with the two approaches above? In principle nothing, but we can try to do more than getting out what we put in. Indeed we are just solving the issue of neutrino masses *ad hoc*, without correlations to other phenomena. In addition we do not improve unification of minimal  $SU(5)$ .

Guided by this double issue of the Georgi-Glashow model, two minimal extensions which can cure at the same time both neutrino masses and unification have been proposed:

- Add a  $15_H = (1, 3)_H \oplus (6, 1)_H \oplus (3, 2)_H$ . Here  $(1, 3)_H$  is an Higgs triplet responsible for type-II seesaw. The model predicts generically light leptoquarks  $(3, 2)_H$  and fast proton decay.
- Add a  $24_F = (1, 1)_F \oplus (1, 3)_F \oplus (8, 1)_F \oplus (3, 2)_F \oplus (\bar{3}, 2)_F$ . Here  $(1, 1)_F$  and  $(1, 3)_F$  are fields responsible respectively for type-I and type-III seesaw. The model predicts a light fermion triplet  $(1, 3)_F$  and fast proton decay.

## 38.5 SUSY Extension

Another well motivated and studied extension of the Georgi-Glashow model is given by SUSY  $SU(5)$ . In this case the supersymmetrization of the spectrum is enough in order to fix both unification and neutrino masses. Indeed,

if we do not impose by hand  $R$ -parity conservation Majorana neutrino masses are automatically generated by lepton number violating interactions.

Consider the particle content of the MSSM. There are three copies of quark and lepton superfields:

$$\begin{aligned} (u, d) &\sim (3, 2, \frac{1}{6}) , & u^c &\sim (\bar{3}, 1, -\frac{2}{3}) , & d^c &\sim (\bar{3}, 1, \frac{1}{3}) , \\ (\nu, e) &\sim (1, 2, -\frac{1}{2}) , & e^c &\sim (1, 1, 1) , \end{aligned} \quad (38.70)$$

and one copy of the two Higgs superfields:

$$(\phi_1^0, \phi_1^-) \sim \left(1, 2, -\frac{1}{2}\right) , \quad (\phi_2^+, \phi_2^0) \sim \left(1, 2, \frac{1}{2}\right) . \quad (38.71)$$

The quarks and leptons can be embedded into  $SU(5)$  as follows:

$$\begin{aligned} \bar{5} &= \left(\bar{3}, 1, \frac{1}{3}\right) + \left(1, 2, -\frac{1}{2}\right) , \\ 10 &= \left(3, 2, \frac{1}{6}\right) + \left(\bar{3}, 1, -\frac{2}{3}\right) + (1, 1, 1) , \end{aligned} \quad (38.72)$$

but the Higgs superfields do not form complete multiplets:  $\Phi_1 \subset \bar{5}$  and  $\Phi_2 \subset 5$ . Their missing partners are  $(\bar{3}, 1, 1/3)$  and  $(3, 1, -1/3)$  respectively and they mediate proton decay. In the MSSM, such effective operators are dimension-five, i.e. they are suppressed by only one power of  $M_U$  in the denominator and can easily contribute to a proton decay lifetime below the experimental lower bound.

Recalling that there is a small discrepancy in the unification of gauge couplings. This can be fixed by threshold corrections due to heavy particles at  $M_U$ . Using these heavy color triplet Higgs superfields to obtain exact unification, it was shown that their masses must lie in the range  $3.5 \times 10^{14}$  GeV to  $3.6 \times 10^{15}$  GeV. However, the experimental lower bound on the decay lifetime of  $p \rightarrow K^+ \bar{\nu}$  is  $6.7 \times 10^{32}$  years, which requires this mass to be greater than  $7.6 \times 10^{16}$  GeV. This contradiction is then used to rule out minimal SUSY  $SU(5)$  as a candidate model of GUT.

The above analysis assumes that the sparticle mass matrices are related to the particle mass matrices in a simple natural way. However, proton decay in the MSSM through the above-mentioned dimension-five operators depends on how sparticles turn into particles. It has been pointed out that if the most general sparticle mass matrices are used, these operators may be sufficiently suppressed to avoid any contradiction with proton decay.

Instead of adjusting the color triplet masses to obtain exact unification, a new and popular way is to invoke extra space dimensions. For example, in a five-dimensional theory, if Higgs fields exist in the bulk, then there can be finite threshold corrections from summing over Kaluza-Klein modes. A specific successful  $SU(5)$  model was proposed using the Kawamura mechanism of symmetry breaking by boundary conditions.

**Exercise 38.1:** Write infinitesimal  $SU(5)$  transformations and covariant derivatives for a 5-dimensional vector and its complex conjugate. Construct the field strength tensor and transformation laws for corresponding gauge bosons.

**Exercise 38.2:** Write down the potential for the scalar fields in the vector representation of  $SU(5)$ . Work out the possible pattern for SSB, how many Goldstone bosons are there in this case? Discuss the possible SSB pattern for the case where there are two such scalar fields.

**Exercise 38.3:** Write down the scalar potential for a scalar fields in the adjoint representation of the  $SU(5)$  group. Work out the possible pattern for the SSB for this field.

**Exercise 38.4:** Show that the  $SU(5)$  antisymmetric tensor representation 10 has the following decomposition

$$10 = (\bar{3}, 1) + (3, 2) + (1, 1),$$

and the adjoint representation 24 has

$$24 = (8, 1) + (1, 3) + (1, 1) + (3, 2) + (\bar{3}, 2).$$

Also, find the decomposition of the symmetric tensor representation 15.

**Exercise 38.5:** If we neglect Higgs in representation 5, we can write the potential by the Higgs field in the adjoint representation of  $SU(5)$ , which is represented as a  $5 \times 5$  hermitian traceless matrix. Show that at the minimum of Higgs potential its diagonal elements can take at most three different values. From this result, discuss the most general form of symmetry breakings that can be induced by a 24 adjoint Higgs field.

**Exercise 38.6:** For the adjoint representation of the Higgs,  $H$ , written as a  $5 \times 5$  traceless hermitian matrix, construct the covariant derivative  $D_\nu H$ . Calculate the mass spectra of the gauge bosons from the covariant derivative if the vacuum expectation value is given by  $\langle H \rangle = v \text{diag}(2, 2, 2, -3, -3)$ .

## Chapter 39

# $SO(10)$ GUT

The next simple Lie group which contains the standard model is

$$SO(10) \supset SU(5) \supset SU(3) \times SU(2) \times U(1) . \quad (39.1)$$

Here, the unification of matter is even more complete, since the irreducible spinor representation 16 contains both the 5 and 10 of  $SU(5)$  and a right-handed neutrino, and thus the complete particle content of one generation of the extended SM with neutrino masses. This is already the largest simple group which achieves the unification of matter in a scheme involving only the already known matter particles (apart from the Higgs sector).

Since different SM fermions are grouped together in larger representations, GUTs specifically predict relations among the fermion masses, such as between the electron and the down quark, the muon and the strange quark, and the tau lepton and the bottom quark. Some of these mass relations hold approximately, but most don't.

The boson matrix for  $SO(10)$  is found by taking the  $15 \times 15$  matrix from the  $10 + 5$  representation of  $SU(5)$  and adding an extra row and column for the right-handed neutrino. The bosons are found by adding a partner to each of the 20 charged bosons (2 right-handed  $W$  bosons, 6 massive charged gluons and 12  $X/Y$  type bosons) and adding an extra heavy neutral  $Z$ -boson to make 5 neutral bosons in total. The boson matrix will have a boson or its new partner in each row and column. These pairs combine to create the familiar 16-dimensional Dirac spinor matrices of  $SO(10)$ .

The power of  $SO(10)$  is historically well-known. A single spinor representation 16 contains the  $\bar{5}$  and 1 of  $SU(5)$  as well as a singlet  $N$ , which may be identified as the right-handed neutrino. The existence of three heavy singlets allows the three known neutrinos to acquire naturally small Majorana masses through the famous seesaw mechanism, and the decay of the lightest of them may also generate a lepton asymmetry in the early Universe which gets converted by sphalerons during the EW phase transition to the present observed baryon asymmetry of the Universe.

What is new in the past years is the realization of the importance of the EW Higgs triplet contained in the 126 of  $SO(10)$ . Whereas the Higgs triplet under  $SU(2)_R$  provides  $N$  with a heavy Majorana mass, the Higgs triplet under  $SU(2)_L$  provides neutrinos with a small Majorana mass. This latter mechanism is also seesaw in character and may in fact be the dominant contribution to the observed neutrino mass.

We shall give the physical foundations of  $SO(10)$  as a GUT group, starting from the SM and browsing in a constructive way through the Georgi-Glashow  $SU(5)$  and the left-right symmetric groups such as the Pati-Salam one. This will offer us the opportunity to introduce the fundamental concepts of GUTs, as charge quantization, gauge unification, proton decay and the connection with neutrino masses in a simplified way.

The  $SO(10)$  gauge group as a candidate for the unification of the elementary interactions was proposed long ago by Georgi and Fritzsche and Minkowski. The main advantage of  $SO(10)$  with respect to  $SU(5)$  is that all the known SM fermions plus three right handed neutrinos fit into three copies of the 16-dimensional spinorial representation of  $SO(10)$ .

In recent years the field received an extra boost due to the discovery of neutrino masses. While in the SM (and in  $SU(5)$ ) there is no rationale for the origin of the small neutrino mass scale, the appeal of  $SO(10)$  consists in the predictive connection between the local  $B - L$  breaking scale (constrained by gauge coupling unification somewhat below  $10^{16}$  GeV) and neutrino masses around 25 orders of magnitude below. Through the implementation of some variant of the seesaw mechanism the inner structure of  $SO(10)$  and its breaking makes natural the appearance of a small neutrino mass scale. This striking connection with neutrino masses is one of the strongest motivations behind  $SO(10)$  and it can be traced back to the left-right symmetric theories which provide a direct connection of the smallness of neutrino masses with the non-observation of the  $V + A$  interactions.



### 39.1 $SO(10)$ Group

$SO(10)$  is the special orthogonal group of rotations in a 10-dimensional vector space. Here special means  $\det O = +1$  which selects the group of transformations continuously connected with the identity. Representation of  $SO(10)$  is given by the group of matrices  $O$  which leave invariant the norm of a 10-dimensional real vector  $\phi$ . Under  $O$ ,  $\phi \rightarrow O\phi$  and since the norm  $\phi^T\phi$  is invariant  $O$  must be orthogonal,  $OO^T = 1$ .

The matrices  $O$  may be written in terms of 45 imaginary generators  $T_{ij} = -T_{ji}$ , for  $i, j = 1, \dots, 10$ , as

$$O = e^{\epsilon_{ij}T_{ij}/2} , \quad (39.2)$$

where  $\epsilon_{ij}$  are the parameters of the transformation. A convenient basis for the generators is

$$(T_{ij})_{ab} = -i (\delta_{a[i}\delta_{b]j}) , \quad (a, b, i, j = 1, \dots, 10) \quad (39.3)$$

where the square bracket stands for anti-symmetrization. They satisfy the  $SO(10)$  commutation relations,

$$[T_{ij}, T_{kl}] = i(\delta_{ik}T_{jl} + \delta_{jl}T_{ik} - \delta_{il}T_{jk} - \delta_{jk}T_{il}) . \quad (39.4)$$

These are an higher dimensional generalization of the well-known  $SO(3)$  commutation relations

$$[J_1, J_2] = i J_3 , \quad \text{where } J_i \equiv \epsilon_{ijk}T_{jk} . \quad (i, j, k = 1, 2, 3) \quad (39.5)$$

Then the right hand side of (39.4) takes just into account the antisymmetric nature of  $T_{ij}$  and  $T_{kl}$ .

In order to study the group theory of  $SO(10)$  it is crucial to identify the invariant tensors. The conditions  $OO^T = 1$  and  $\det O = +1$  give rise to two of them.

- The first one is simply the Kronecker tensor  $\delta_{ij}$  which is easily proven to be invariant because of  $OO^T = 1$ , namely

$$\delta_{ij} \rightarrow O_{ik}O_{jl}\delta_{kl} = O_{ik}O_{jk} = \delta_{ij} . \quad (39.6)$$

- The second one is the 10-index Levi-Civita tensor  $\epsilon_{ijklmnopqr}$ . Indeed, from the definition of determinant

$$\begin{aligned} & \det O \epsilon_{i'j'k'l'm'n'o'p'q'r'} = \\ & = O_{i'i} O_{j'j} O_{k'k} O_{l'l} O_{m'm} O_{n'n} O_{o'o} O_{p'p} O_{q'q} O_{r'r} \epsilon_{ijklmnopqr} \end{aligned} \quad (39.7)$$

and the fact that  $\det O = +1$ , we conclude that  $\epsilon_{ijklmnopqr}$  is also invariant.

The irreducible representations of  $SO(10)$  can be classified into two categories, single and double valued representations. The single-valued representations have the same transformations properties as the ordinary vectors in the real 10-dimensional space and their symmetrized or antisymmetrized tensor products. The double-valued representations, called also spinor representations, transform like spinors in a 10-dimensional coordinate space.

### 39.1.1 Tensor Representations

The general  $n$ -index irreducible representations of  $SO(10)$  are built by means of the antisymmetrization or symmetrization (including trace subtraction) of the tensor product of  $n$ -fundamental vectors. Starting from the 10-dimensional fundamental vector  $\phi_i$ , whose transformation rule is

$$\phi_i \rightarrow O_{ij} \phi_j, \quad (39.8)$$

we can decompose the tensor product of two of them in the following way

$$\begin{aligned} \phi_i \times \phi_j = & \underbrace{(\phi_i \times \phi_j - \phi_j \times \phi_i) / 2}_{\phi_{ij}^A} + \underbrace{\delta_{ij} \phi_k \times \phi_k / 10}_{S\delta_{ij}} + \\ & + \underbrace{(\phi_i \times \phi_j + \phi_j \times \phi_i) / 2 - \delta_{ij} \phi_k \times \phi_k / 10}_{\phi_{ij}^S}. \end{aligned} \quad (39.9)$$

Since the symmetry properties of tensors under permutation of the indices are not changed by the group transformations, the antisymmetric tensor  $\phi_{ij}^A$  and the symmetric tensor  $\phi_{ij}^S$  clearly do not transform into each other.

In general one can also separate a tensor in a traceless part and a trace. Because  $O$  is orthogonal also the traceless property is preserved by the group transformations. So we conclude that  $\phi_{ij}^A$ ,  $\phi_{ij}^S$  and  $S\delta_{ij}$  form irreducible

representations whose dimensions are respectively  $10(10-1)/2 = 45$ ,  $10(10+1)/2 - 1 = 54$  and 1. One can continue in this way by considering higher order representations and separating each time the symmetric/antisymmetric pieces and subtracting traces.

Something special happens for 5-index tensors and the reason has to do with the existence of the invariant  $\epsilon_{ijklmnopqr}$  which induces the following duality map when applied to a 5-index completely antisymmetric tensor  $\phi_{nopqr}$ ,

$$\phi_{ijklm} \rightarrow \tilde{\phi}_{ijklm} \equiv -\frac{i}{5!} \epsilon_{ijklmnopqr} \phi_{nopqr} . \quad (39.10)$$

This allows us to define the self-dual and the antiself-dual components of  $\phi_{ijklm}$  in the following way

$$\Sigma_{ijklm} \equiv \frac{1}{\sqrt{2}} \left( \phi_{ijklm} + \tilde{\phi}_{ijklm} \right) , \quad (39.11)$$

$$\bar{\Sigma}_{ijklm} \equiv \frac{1}{\sqrt{2}} \left( \phi_{ijklm} - \tilde{\phi}_{ijklm} \right) . \quad (39.12)$$

One verifies that  $\tilde{\Sigma}_{ijklm} = \Sigma_{ijklm}$  (self-dual) and  $\tilde{\bar{\Sigma}}_{ijklm} = -\bar{\Sigma}_{ijklm}$  (antiself-dual). Since the duality property is not changed by the group transformations,  $\Sigma_{ijklm}$  and  $\bar{\Sigma}_{ijklm}$  form irreducible representations with the dimension

$$\frac{1}{2} \frac{10!}{5!(10-5)!} = 126 . \quad (39.13)$$

### 39.1.2 Spinor Representations

We have defined the  $SO(10)$  group by those linear transformations on the coordinates  $x_1, x_2, \dots, x_{10}$ , such that the quadratic form  $x_1^2 + x_2^2 + \dots + x_{10}^2$  is left invariant. If we write this quadratic form as the square of a linear form of  $x_i$ 's,

$$x_1^2 + x_2^2 + \dots + x_{10}^2 = (\gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_{10} x_{10})^2 , \quad (39.14)$$

we have to require

$$\{\gamma_i, \gamma_j\} = 2\delta_{ij} . \quad (39.15)$$

This goes under the name of Clifford algebra and the  $\gamma$ 's have to be matrices in order to anticommute with each other. In particular it can be shown that the dimension of the  $\gamma$  matrices must be even. Indeed from (39.15) we obtain

$$\gamma_j(\gamma_i \gamma_j + \gamma_j \gamma_i) = 2\gamma_j \quad \text{or} \quad \gamma_j \gamma_i \gamma_j = \gamma_i , \quad (39.16)$$

with no sum over  $j$ . Taking the trace we get

$$\text{Tr}\gamma_j\gamma_i\gamma_j = \text{Tr}\gamma_i . \quad (39.17)$$

But for the case  $i \neq j$  this implies

$$\text{Tr}\gamma_j\gamma_i\gamma_j = -\text{Tr}\gamma_i\gamma_j\gamma_j = -\text{Tr}\gamma_i . \quad (39.18)$$

Putting together (39.17) and (39.18), we have  $\text{Tr}\gamma_i = 0$ .

On the other hand,  $\gamma_i^2 = 1$  implies that the eigenvalues of  $\gamma_i$  are either  $+1$  or  $-1$ . This means that to get  $\text{Tr}\gamma_i = 0$ , the number of  $+1$  and  $-1$  eigenvalues must be the same, i.e.  $\gamma_i$  must be even dimensional.

For definiteness let us build an explicit representation of the  $\gamma$ 's which is valid for  $SO(2N)$  groups. We start with  $N = 1$ . Since the Pauli matrices satisfy the Clifford algebra

$$\{\sigma_i, \sigma_j\} = 2\delta_{ij} , \quad (39.19)$$

we can choose

$$\gamma_1^{(1)} = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \gamma_2^{(1)} = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} . \quad (39.20)$$

Then the case  $N > 1$  is constructed by recursion. The iteration from  $N$  to  $N + 1$  is defined by

$$\gamma_i^{(N+1)} = \begin{pmatrix} \gamma_i^{(N)} & 0 \\ 0 & -\gamma_i^{(N)} \end{pmatrix} , \quad (i = 1, 2, \dots, 2N) \quad (39.21)$$

with

$$\gamma_{2N+1}^{(N+1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \gamma_{2N+2}^{(N+1)} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} . \quad (39.22)$$

Given the fact that the  $\gamma_i^{(N)}$  matrices satisfy the Clifford algebra let us check

explicitly that the  $\gamma_i^{(N+1)}$  ones satisfy it as well,

$$\begin{aligned} \left\{ \gamma_i^{(N+1)}, \gamma_j^{(N+1)} \right\} &= \begin{pmatrix} \left\{ \gamma_i^{(N)}, \gamma_j^{(N)} \right\} & 0 \\ 0 & \left\{ \gamma_j^{(N)}, \gamma_i^{(N)} \right\} \end{pmatrix} = \\ &= \begin{pmatrix} 2\delta_{ij} & 0 \\ 0 & 2\delta_{ij} \end{pmatrix} = 2\delta_{ij} , & (39.23) \\ \left\{ \gamma_i^{(N+1)}, \gamma_{2N+1}^{(N+1)} \right\} &= \begin{pmatrix} 0 & \gamma_i^{(N)} \\ -\gamma_i^{(N)} & 0 \end{pmatrix} + \begin{pmatrix} 0 & -\gamma_i^{(N)} \\ \gamma_i^{(N)} & 0 \end{pmatrix} = 0 , \\ \left( \gamma_{2N+1}^{(N+1)} \right)^2 &= 1 . \end{aligned}$$

Analogously one finds

$$\begin{aligned} \left\{ \gamma_i^{(N+1)}, \gamma_{2N+2}^{(N+1)} \right\} &= 2\delta_{ij} , \\ \left\{ \gamma_{2N+1}^{(N+1)}, \gamma_{2N+2}^{(N+1)} \right\} &= 0 , & (39.24) \\ \left( \gamma_{2N+2}^{(N+1)} \right)^2 &= 1 . \end{aligned}$$

Now consider a rotation in the coordinate space,  $x'_i = O_{ik}x_k$ , where  $O$  is an orthogonal matrix. This rotation induces a transformation on the  $\gamma_i$  matrix

$$\gamma'_i = O_{ik}\gamma_k . \quad (39.25)$$

Notice that the anticommutation relations remain unchanged, i.e.

$$\{\gamma'_i, \gamma'_j\} = O_{ik}O_{jl}\{\gamma_k, \gamma_l\} = 2\delta_{ij} . \quad (39.26)$$

Because the original set of  $\gamma$  matrices form a complete matrix algebra, the new set of  $\gamma$  matrices must be related to the original set by a similarity transformation,

$$\gamma'_i = S(O)\gamma_i S^{-1}(O) \quad \text{or} \quad O_{ik}\gamma_k = S(O)\gamma_i S^{-1}(O) . \quad (39.27)$$

The correspondence  $O \rightarrow S(O)$  serves as a  $2^N$ -dimensional representation of the rotation group which is called spinor representation. The quantities  $\psi_i$ , which transform like

$$\psi'_i = S(O)_{ij}\psi_j , \quad (39.28)$$

are called spinors. For an infinitesimal rotation we can parametrize  $O_{ik}$  and  $S(O)$  by

$$O_{ik} = \delta_{ik} + \epsilon_{ik} \quad \text{and} \quad S(O) = 1 + \frac{i}{2}S_{ij}\epsilon_{ij} , \quad (39.29)$$

with  $\epsilon_{ik} = -\epsilon_{ki}$ . Then (39.27) implies

$$i[S_{kl}, \gamma_i] = (\gamma_l \delta_{ik} - \gamma_k \delta_{il}) , \quad (39.30)$$

where we have used

$$\epsilon_{ik} \gamma_k = \epsilon_{lk} \gamma_k \delta_{il} = \frac{1}{2} (\gamma_k \delta_{il} - \gamma_k \delta_{jl}) . \quad (39.31)$$

One can verify that a solution for  $S_{kl}$  in (39.30) is

$$S_{kl} = \frac{i}{4} [\gamma_k, \gamma_l] . \quad (39.32)$$

By expressing the parameter  $\epsilon_{kl}$  in terms of rotations angle, one can see that  $S(O(4\pi)) = 1$ , i.e.  $S(O)$  is a double-valued representation.

This is easily seen for  $SO(3)$ . In this case the Clifford algebra is simply given by the three Pauli matrices and a finite transformation looks like

$$S(O(\varphi)) = e^{\frac{i}{2} \sigma_i \varphi_i} = \cos \frac{|\varphi|}{2} + i \frac{\sigma_i \varphi_i}{|\varphi|} \sin \frac{|\varphi|}{2} , \quad (39.33)$$

where we have defined

$$\epsilon_{23} \equiv -\varphi_1 , \quad \epsilon_{13} \equiv -\varphi_2 , \quad \epsilon_{12} \equiv -\varphi_3 , \quad |\varphi| = \sqrt{\varphi_1^2 + \varphi_2^2 + \varphi_3^2} . \quad (39.34)$$

However for  $SO(2N)$  groups the representation  $S(O)$  is not irreducible. To see this we construct the chiral projector  $\gamma_\chi$  defined by

$$\gamma_\chi = (-i)^N \gamma_1 \gamma_2 \cdots \gamma_{2N} . \quad (39.35)$$

$\gamma_\chi$  anticommutes with  $\gamma_i$  since  $2N$  is even and consequently we get  $[\gamma_\chi, S_{kl}] = 0$ . Thus if  $\psi$  transforms as  $\psi'_i = S(O)_{ij} \psi_j$ , the positive and negative chiral components

$$\psi^+ \equiv \frac{1}{2} (1 + \gamma_\chi) \psi \quad \text{and} \quad \psi^- \equiv \frac{1}{2} (1 - \gamma_\chi) \psi \quad (39.36)$$

transform separately. In other words  $\psi^+$  and  $\psi^-$  form two irreducible spinor representations of dimension  $2^{N-1}$ .

To find the relation between  $\psi^+$  and  $\psi^-$  it is necessary to introduce the concept of conjugation. Let us consider a spinor  $\psi$  of  $SO(2N)$ . The combination  $\psi^T C \psi$  is an  $SO(2N)$  invariant provided that

$$S_{ij}^T C = -C S_{ij} . \quad (39.37)$$

The conjugation matrix  $C$  can be constructed iteratively. We start from  $C^{(1)} = i\sigma_2$  for  $N = 1$  and define

$$C^{(N+1)} = \begin{pmatrix} 0 & C^{(N)} \\ (-)^{(N+1)}C^{(N)} & 0 \end{pmatrix}. \quad (39.38)$$

One can verify that

$$(C^{(N)})^{-1}\gamma_i^T C^{(N)} = (-)^N \gamma_i. \quad (39.39)$$

By transposing (39.39) and substituting back  $\gamma_i^T$  we get

$$[\gamma_i, ((C^{(N)})^T)^{-1}C^{(N)}] = 0. \quad (39.40)$$

Then the Shur's Lemma implies

$$((C^{(N)})^T)^{-1}C^{(N)} = \lambda I \quad \text{or} \quad C^{(N)} = \lambda(C^{(N)})^T, \quad (39.41)$$

which yields  $\lambda^2 = 1$ . In order to choose between  $\lambda = +1$  and  $\lambda = -1$  one has to apply (39.38), obtaining

$$C^T = (-)^{N(N+1)/2} C. \quad (39.42)$$

On the other hand (39.35) and (39.39) lead to

$$(C^{(N)})^{-1}\gamma_\chi^T C^{(N)} = (-)^N \gamma_\chi, \quad (39.43)$$

which by exploiting  $\gamma_\chi^T = \gamma_\chi$  yields

$$(C^{(N)})^{-1}\gamma_\chi C^{(N)} = (-)^N \gamma_\chi. \quad (39.44)$$

This allows us to write

$$\begin{aligned} (C^{(N)})^{-1} (S_{ij}(1 + \gamma_\chi))^* C^{(N)} &= (C^{(N)})^{-1} S_{ij}^* (1 + \gamma_\chi) C^{(N)} = \\ &- S_{ij} (1 + (-)^N \gamma_\chi). \end{aligned} \quad (39.45)$$

where we have also exploited the hermicity of the  $\gamma$  matrices. (39.45) can be interpreted in the following way: for  $SO(2N)$  with  $N$  even  $\psi^+$  and  $\psi^-$  are self-conjugate, i.e. real or pseudo-real depending on whether  $C$  is symmetric or antisymmetric, while for  $SO(2N)$  with  $N$  odd  $\psi^+$  is the conjugate of  $\psi^-$ . Thus only  $SO(4k + 2)$  can have complex representations and remarkably  $SO(10)$  belong to this class.

Note a distinctive feature of spinorial representations: spinors of  $SO(2N)$  decompose into the direct sum of spinors of  $SO(2N') \subset SO(2N)$ . Indeed, since the construction of  $\gamma_\chi$  in (39.35) is such that

$$\gamma_\chi^{(N+1)} = \begin{pmatrix} \gamma_\chi^{(N)} & 0 \\ 0 & -\gamma_\chi^{(N)} \end{pmatrix}, \quad (39.46)$$

the positive-chirality spinor  $\psi^+$  of  $SO(2N + 2M)$  contains  $2^{M-1}$  positive-chirality spinors and  $2^{M-1}$  negative-chirality spinors of  $SO(2N)$ . More explicitly

$$\begin{aligned} \psi_{SO(2N+2M)}^+ &\rightarrow \psi_{SO(2N+2M-2)}^+ \oplus \psi_{SO(2N+2M-2)}^- \\ &\rightarrow 2 \times \psi_{SO(2N+2M-4)}^+ \oplus 2 \times \psi_{SO(2N+2M-4)}^- \rightarrow \cdots \\ &\rightarrow 2^{M-1} \times \psi_{SO(2N)}^+ \oplus 2^{M-1} \times \psi_{SO(2N)}^-. \end{aligned} \quad (39.47)$$

Let us exemplify this important concept in the case of the 16-dimensional positive-chirality spinor of  $SO(10)$ . By taking respectively  $(N = 3, M = 2)$  and  $(N = 2, M = 3)$  we obtain

- $16 = 2 \times 4^+ \oplus 2 \times 4^-$  under  $SO(10) \supset SO(6)$ ,
- $16 = 4 \times 2^+ \oplus 4 \times 2^-$  under  $SO(10) \supset SO(4)$ ,

where  $4^+$  ( $4^-$ ) and  $2^+$  ( $2^-$ ) are respectively the positive (negative) chiral components of the  $SO(6)$  and  $SO(4)$  reducible spinors. Thus under  $SO(10) \supset SO(6) \times SO(4)$  the 16 decomposes as

$$16 = (4^+, 2^+) \oplus (4^-, 2^-). \quad (39.48)$$

The Lie algebras  $SO(6)$  and  $SO(4)$  are isomorphic to  $SU(4)$  and  $SU(2) \times SU(2)$ . This allows us to make the following identifications between the  $SO(6)$  and  $SU(4)$  representations

$$4^+ \sim 4, \quad 4^- \sim \bar{4}, \quad (39.49)$$

and the  $SO(4)$  and  $SU(2) \times SU(2)$  ones

$$2^+ \sim (2, 1), \quad 2^- \sim (1, 2), \quad (39.50)$$



which justify the decomposition of the  $SO(10)$  spinor under the Pati-Salam algebra  $SU(4)_C \times SU(2)_L \times SU(2)_R$ , namely

$$16 = (4, 2, 1) \oplus (\bar{4}, 1, 2) . \quad (39.51)$$

This striking group-theoretic feature of spinors, which under the natural restriction to an orthogonal subgroup decompose into several copies of identical spinors of the subgroup, hints to a suggestive connection with the repetitive structure of the SM families and motivates the study of unification in higher orthogonal groups than  $SO(10)$ . To accommodate at least the three observed matter families we must use either  $SO(16)$  or  $SO(18)$ . Following the decomposition in (39.47) we get

- $SO(16)$ :  $\psi_{SO(16)}^+ \rightarrow 4 \times \psi_{SO(10)}^+ \oplus 4 \times \psi_{SO(10)}^-$  ,
- $SO(18)$ :  $\psi_{SO(18)}^+ \rightarrow 8 \times \psi_{SO(10)}^+ \oplus 8 \times \psi_{SO(10)}^-$  .

However there is a fundamental difference between the two cases above.

According to the discussion below (39.45) only  $SO(4k + 2)$  groups have complex spinor representations. This means that one can write a super-heavy bare mass term for  $\psi_{SO(16)}^+$  and it is difficult to explain why it should be light. On the other hand no bare mass term can be written for  $\psi_{SO(18)}^+$ , making the last group a more natural choice.

The obvious difficulty one encounters in this class of models is the overabundance of sequential or mirror families. If we decide to embed the SM fermions into three copies of  $\psi_{SO(10)}^+$ , the remaining families in  $\psi_{SO(10)}^+$  are called sequential, while those in  $\psi_{SO(10)}^-$  are mirror families.

Mirror fermions have the identical quantum numbers of ordinary fermions under the SM gauge group, except that they have opposite handedness. They imply parity restoration at high-energies as proposed long ago by Lee and Yang. It has been pointed out recently that the existence of three (mirror or sequential) families is still in accord with the SM, as long as an additional Higgs doublet is also present. This however is not enough to allow large orthogonal unification scenarios based on  $SO(16)$  or  $SO(18)$ .

## 39.2 Anomaly Cancellation

$SO(10)$  is an anomaly-free group. This important property can be understood from a simple group theoretical argument.

Let us consider the  $SO(10)$  generators  $T_{ij}$  in a given arbitrary representation.  $T_{ij}$  transforms like an antisymmetric tensor in the indices  $i$  and  $j$ . Then the anomaly, which is proportional to the invariant tensor

$$\text{Tr}\{T_{ij}, T_{kl}\}T_{mn} , \quad (39.52)$$

must be a linear combination of a product of Kronecker  $\delta$ 's.

Furthermore it must be antisymmetric under the exchanges  $i \leftrightarrow j$ ,  $k \leftrightarrow l$ ,  $m \leftrightarrow n$  and symmetric under the exchange of pairs  $ij \leftrightarrow kl$ ,  $kl \leftrightarrow mn$  and  $ij \leftrightarrow mn$ . However the most general form consistent with the antisymmetry in  $i \leftrightarrow j$ ,  $k \leftrightarrow l$ ,  $m \leftrightarrow n$

$$\begin{aligned} & \delta_{jk}\delta_{lm}\delta_{ni} - \delta_{ik}\delta_{lm}\delta_{nj} - \delta_{jl}\delta_{km}\delta_{ni} + \delta_{il}\delta_{km}\delta_{nj} - \\ & - \delta_{jk}\delta_{ln}\delta_{mi} + \delta_{ik}\delta_{ln}\delta_{mj} + \delta_{jl}\delta_{kn}\delta_{mi} - \delta_{il}\delta_{kn}\delta_{mj} , \end{aligned} \quad (39.53)$$

is antisymmetric in  $ij \leftrightarrow kl$  as well and so it must vanish. The proof fails for  $SO(6)$  where the anomaly can be proportional to the invariant tensor  $\epsilon_{ijklmn}$ . Actually this is consistent with the fact that  $SO(6)$  is isomorphic to  $SU(4)$  which is clearly an anomalous group. On the other hand  $SO(N)$  is safe for  $N > 6$ .

## 39.3 The SM Embedding

From the  $SO(10)$  commutation relations in (39.4) we find that a complete set of simultaneously commuting generators can be chosen as

$$T_{12} , T_{34} , T_{56} , T_{78} , T_{90} . \quad (39.54)$$

This is also known as the Cartan subalgebra and can be spanned over the left-right group Cartan generators

$$T_C^3 , T_C^8 , T_L^3 , T_R^3 , T_{B-L} . \quad (39.55)$$

Let us consider the  $SO(4) \times SO(6)$  maximal subalgebra of  $SO(10)$ . We can span the  $SO(4)$  generators over  $T_{ij}$  with  $i, j = 1, 2, 3, 4$  and the  $SO(6)$

generators over  $T_{ij}$  with  $i, j = 5, 6, 7, 8, 9, 10$ . From the  $SO(10)$  commutation relations in (39.4) one can verify that these two sets commute, hence the direct product  $SO(4) \times SO(6)$ .

The next information we need is the notion of local isomorphism for the algebras

$$SO(4) \sim SU(2) \times SU(2), \quad SO(6) \sim SU(4). \quad (39.56)$$

In the  $SO(4)$  case we define

$$T_{L,R}^1 \equiv \frac{T_{23} \pm T_{14}}{2}, \quad T_{L,R}^2 \equiv \frac{T_{31} \pm T_{24}}{2}, \quad T_{L,R}^3 \equiv \frac{T_{12} \pm T_{34}}{2}, \quad (39.57)$$

and check by an explicit calculation that

$$[T_L^i, T_L^j] = i \epsilon^{ijk} T_L^k, \quad [T_R^i, T_R^j] = i \epsilon^{ijk} T_R^k, \quad [T_L^i, T_R^j] = 0. \quad (39.58)$$

Thus  $T_L^i$  and  $T_R^i$  ( $i = 1, 2, 3$ ) span respectively the  $SU(2)_L$  and the  $SU(2)_R$  algebra. On the other hand for the  $SO(6)$  sector we define

$$\begin{aligned} T_C^1 &\equiv \frac{T_{89} + T_{70}}{2}, & T_C^2 &\equiv \frac{T_{97} + T_{80}}{2}, & T_C^3 &\equiv \frac{T_{09} + T_{87}}{2}, \\ T_C^4 &\equiv \frac{T_{96} + T_{05}}{2}, & T_C^5 &\equiv \frac{T_{59} + T_{06}}{2}, & T_C^6 &\equiv \frac{T_{67} + T_{85}}{2}, \\ T_C^7 &\equiv \frac{T_{75} + T_{86}}{2}, & T_C^8 &\equiv \frac{2T_{65} + T_{78} + T_{09}}{2\sqrt{3}}, & T_C^9 &\equiv \frac{T_{67} + T_{58}}{2}, \\ T_C^{10} &\equiv \frac{T_{75} + T_{68}}{2}, & T_C^{11} &\equiv \frac{T_{69} + T_{05}}{2}, & T_C^{12} &\equiv \frac{T_{95} + T_{06}}{2}, \\ T_C^{13} &\equiv \frac{T_{89} + T_{07}}{2}, & T_C^{14} &\equiv \frac{T_{97} + T_{08}}{2}, & T_C^{15} &\equiv \frac{T_{65} + T_{87} + T_{90}}{\sqrt{6}}, \end{aligned} \quad (39.59)$$

and verify after a tedious calculation that

$$[T_C^i, T_C^j] = i f^{ijk} T_C^k, \quad (39.60)$$

where  $f^{ijk}$  are the structure constants of  $SU(4)$ . Thus  $T_C^i$  ( $i = 1, \dots, 15$ ) spans the  $SU(4)_C$  algebra and, in particular, the  $SU(3)_C$  subalgebra is spanned by  $T_C^i$  ( $i = 1, \dots, 8$ ) while  $T_C^{15}$  can be identified with the (normalized)  $T_{B-L}$  generator. Then the hypercharge and electric charge operators read respectively

$$Y = T_R^3 + \sqrt{\frac{2}{3}} T_{B-L} = \frac{1}{2} (T_{12} - T_{34}) + \frac{1}{3} (T_{65} + T_{87} + T_{90}) \quad (39.61)$$

and

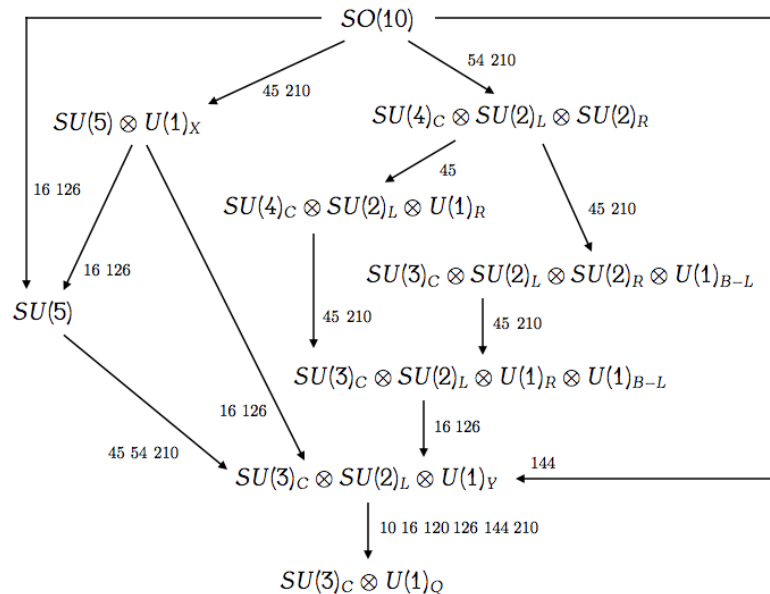
$$Q = T_L^3 + Y = T_{12} + \frac{1}{3} (T_{65} + T_{87} + T_{90}). \quad (39.62)$$

### 39.4 The Higgs Sector

As we have seen above  $SO(10)$  offers a powerful organizing principle for the SM matter content whose quantum numbers nicely fit in a 16-dimensional spinorial representation. However there is an obvious prize to pay: the more one unifies the more one has to work in order to break the enhanced symmetry.

The symmetry breaking sector can be regarded as the most arbitrary and challenging aspect of GUT models. The standard approach is based on the SSB through elementary scalars. Though other ways to face the problem may be conceived the Higgs mechanism remains the most solid one in terms of computability and predictivity. Establishing the minimal Higgs content needed for the GUT breaking is a basic question which has been addressed since the early days of the GUT program.

The figure below shows  $SO(10)$  breaking chart with representations up to the 210.



In this figure it is indicated the possible symmetry stages between  $SO(10)$

and  $SU(3)_C \times U(1)_Q$  with the corresponding scalar representations responsible for the breaking. That gives an idea of the complexity of the Higgs sector in  $SO(10)$  GUTs.  $SU(5) \times U(1)_X$  can be understood either in the standard or in the flipped realization. In the former case 16 or 126 breaks it into  $SU(5)$ , while in the latter into  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . For simplicity we are neglecting the distinctions due to the discrete left-right symmetry.

In view of such a degree of complexity, better we start by considering a minimal Higgs sector. Let us stress that the quest for the simplest Higgs sector is driven not only by aesthetic criteria but it is also a phenomenologically relevant issue related to tractability and predictivity of the models.

Indeed, the details of the symmetry breaking pattern, sometimes overlooked in the phenomenological analysis, give further constraints on the low-energy observables such as the proton decay and the effective SM flavor structure. For instance in order to assess quantitatively the constraints imposed by gauge coupling unification on the mass of the lepto-quarks responsible for proton decay it is crucial to have the scalar spectrum under control.

From the breaking chart we conclude that, before considering any symmetry breaking dynamics, the following representations are required by the group theory in order to achieve a full breaking of  $SO(10)$  down to the SM:

- $16_H$  or  $126_H$ : they reduce the rank by one unit but leave an  $SU(5)$  little group unbroken;
- $45_H$  or  $54_H$  or  $210_H$ : they admit for little groups different from  $SU(5) \otimes U(1)$ , yielding the SM when intersected with  $SU(5)$ .

It should be also mentioned that a one-step  $SO(10) \rightarrow$  SM breaking can be achieved via only one  $144_H$  irreducible Higgs representation. However, such a setting requires an extended matter sector, including  $45_F$  and  $120_F$  multiplets, in order to accommodate realistic fermion masses.

The dynamics of the SSB imposes further constraints on the viability of the options showed in the chart. On top of that one has to take into account also other phenomenological constraints due to the unification pattern, the proton decay and the SM fermion spectrum. While the choice between  $16_H$  or  $126_H$  is a model dependent issue related to the details of the Yukawa sector, the simplest option among  $45_H$ ,  $54_H$  and  $210_H$  is certainly given by the adjoint  $45_H$ .

However, since the early 80's, it has been observed that the vacuum dynamics aligns the adjoint along an  $SU(5) \times U(1)$  direction, making the choice of  $16_H$  (or  $126_H$ ) and  $45_H$  alone not phenomenologically viable. In the non-SUSY case the alignment is only approximate. but it is such to clash with unification constraints which do not allow for any  $SU(5)$ -like intermediate stage, while in the SUSY limit the alignment is exact due to  $F$ -flatness, thus never landing to a supersymmetric SM vacuum.

**Exercise 39.1:** Write infinitesimal  $SO(10)$  transformations for two sets of scalar fields in the vector representations of  $SO(10)$  and construct the covariant derivative for one of them.

**Exercise 39.2:** Show that in the group  $SO(n)$  with either even  $n = 2m$  or odd  $n = 2m + 1$ , we can find  $m$  mutually commuting generators.

**Exercise 39.3:** Consider the  $n$ -dimensional real space. Show that if we write the quadratic form of this space as the square of a linear form, then the coefficient of the coordinates in this expression satisfy the anticommutation relation, which is usually referred to as the Clifford algebra. Show also that if we take this coefficients to be hermitian matrices, then they have to be even-dimensional matrices.

**Exercise 39.4:** Show that the  $SU(5)$  is a subgroup of  $SO(10)$ .

**Exercise 39.5:** Show that under  $SU(5)$ , the spinor representation  $S^+ = 16$  of  $SO(10)$  reduces as  $16 = 10 + 5 + 1$ .

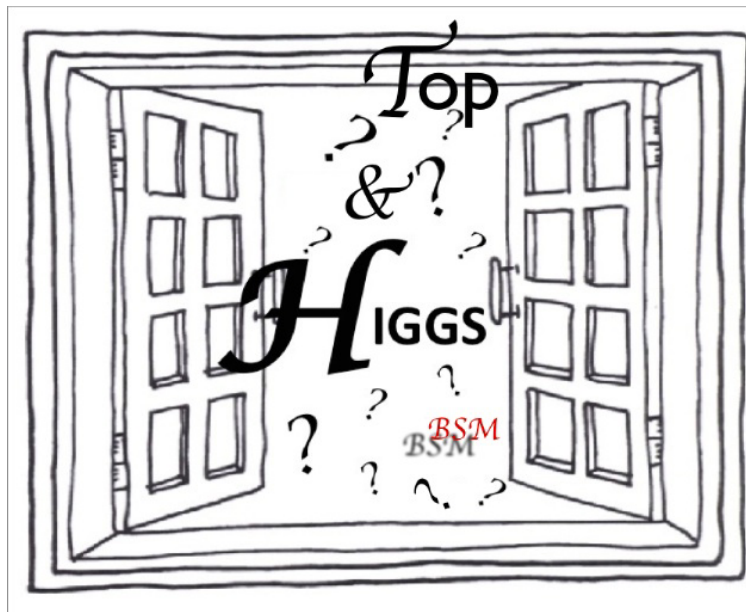
## Part XIV

# Afterwards





In any case the days of SM are coming to an end in some sense! Hopefully it will be the case of "The King is dead" and "Long live the King"! We have, however, not much idea what particular *Beyond the SM* (BSM) option, if any, would be the new king. As we have discussed in these lectures, already the mass of the observed states of SM particles can be used to answer the question about the scale up to which the SM is valid. In fact, this has been one of the most impressive facts about the SM. It has held the ability to ask and answer questions about its own consistency within its structure. Just like the *gauge principle* and the *unitarity* were the guiding principle so far, now the *small* mass of the discovered Higgs ( $\sim \mathcal{O}$  weak scale) might be the guiding principle for future theoretical developments! We should get a peek at the BSM land through the "window" of measurement of the properties of the Higgs and the top quark!



Exciting days are ahead for sure! If LHC will fail to find *direct evidence* for the BSM physics we would really have to understand what is so special about the SM. Precision measurements of the observed Higgs mass and Higgs couplings will be then our window to this world of physics beyond the SM.

*String Theory* or an *Extra Dimension Model*, if true, must be able to reproduce the same general framework we have seen so far. But, String Theory is fundamentally a geometric construct and will reproduce everything we have

seen about gauge theory, but from a geometric framework. This should not be entirely foreign, though. Recall that, for electromagnetism, the gauge group is  $U(1)$ . We can "draw" this geometrically as a circle in the complex plane. The Weak force is represented by the gauge group  $SU(2)$ , which we have seen is parameterized by three numbers, and therefore has three generators. As we discussed in these lectures, we should think of these spaces as vector spaces and the generators as basis vectors spanning the entire space. The same is true of  $SU(3)$ , though it is an eight-dimensional space. So, because there is a space associated with each of these groups, it should be somewhat obvious that there is a natural geometric picture associated with a Lie group.

While the idea of the parameter space of a Lie group having a geometric picture associated with it may seem straightforward, the geometry undergirding gauge theory can be extremely complicated, and we therefore must spend a significant amount of time investigating it. Just as we have built gauge theory from algebra, we will in a sense start over and rebuild it using geometry.

When we finally get to String and Extra Dimensional Theories, we will see that the geometric and algebraic pictures come together beautifully, and that a thorough understanding of both will be necessary to understand what may be the "ultimate" theory of our universe.