CERN–2005–007
1 July 2005

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

2003 EUROPEAN SCHOOL OF HIGH-ENERGY PHYSICS

Tsakhkadzor, Armenia
24 August–6 September 2003

PROCEEDINGS

Editor: A. Olchevski

GENEVA
2005

CERN–XXX copies printed–July 2005

ABSTRACT

The European School of High Energy Physics is intended to give young experimental physicists an introduction to the theoretical aspects of recent advances in elementary particle physics. These proceedings contain lecture notes on field theory and the Standard Model, flavour physics and CP violation, quantum chromodynamics, cosmic rays, astroparticle physics, beyond the Standard Model and finally, multiparticle dynamics.

PREFACE

The eleventh in the new series of the European School of High-Energy Physics, which took place in Tsakhkadzor, Armenia, from 24 August to 6 September 2003, was attended by sixty-nine students coming from eighteen different countries (plus four Armenian students invited as listeners). The School was hosted in the 'House of the Writers' a somewhat old-fashioned, but beautiful hotel located in Tsakhkadzor, a small town about 50 km north of Yerevan. The origin of this place goes back to the 11th century when the magister Grigor Pahlavuny founded a church named as Kecharis (later renamed Tsakhkadzor which means 'canyon of the flowers'). Nowadays this complex of four beautiful churches with patterned cross-stones is a pride of Armenian architecture. The little town — a well-known ski resort in the sixties — is located in a valley surrounded by beautiful mountains.

According to the tradition of the School, the students were sharing twin rooms, mixing nationalities and in particular Eastern participants with Western ones. The meals were served on a buffet style, and the food was typically Armenian with a lot of vegetables mixed with meat.

Doctor George Pogosyan, Director of the Physics Department at the Yerevan State University, was acting as local director of the School.

Our sincere thanks go to George and his team of Armenian students who, together with the local committee, made it possible to organize the School and contributed to its success. Our thanks are also due to the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students, who in turn manifested their good spirits during two intense weeks, undoubtedly appreciated their personal contribution in answering questions and explaining points of theory.

We are very grateful to Tatyana Donskova and Danielle Métral for their untiring efforts in the lengthy preparations for and the day-to-day care of the School. Their efficient teamwork and continuous care of the students and their needs were highly appreciated. Our special thanks also go to the hotel manager who, with her background as a nurse, provided first aid when the school was hit by an epidemic of stomach problems. She also 'requisitioned' Doctor Avagyan, who happened to be staying at the hotel, and together they made it possible for the participants to complete the school and return in good shape to their home countries.

The School participants enjoyed several memorable excursions. Of particular interest was the visit to the Matanaderan (Yerevan), the Institute of old medieval manuscripts, including fragments from the 5th and 6th centuries, and where we could see the only existing version of Ibn Batuta's (Avicenna's) book of geometry, which demonstrates that gravity was discussed from early medieval times. The participants also enjoyed the excursion to the cosmic-ray high altitude station on Mt. Aragats: first at the Nor-Ambert station at 2000 metres, where Professor Ashot Chilingarian gave a special lecture, and then at the Aragats station at 3200 metres.

However, the success of the 2003 School was to a large extent due to the students themselves. Their posters displayed on the walls of the beautiful hall of the 'House of the Writers' were of excellent quality both technically and in content, and throughout the School they participated actively during the lectures, in the discussion sessions, and with genuine interest in the different activities and excursions.

Egil Lillestøl
on behalf of the Organizing Committee

PEOPLE IN THE PHOTOGRAPH

Name	Number
Mathieu Agelou	22
Robert Avakian	62
Evgueni Baldine	36
Konstantin Batkov	8
Simon Baumgartner	77
Stéphanie Beauceron	72
Olaf Behrendt	61
Ilija Bizjak	63
Ingo Bloch	44
Carmela Carpentieri	5
Roger Cashmore	65
Antoine Cazes	53
Ashot Chilingarian	38
Alexis Cothenet	57
Laura Covi	43
Peter Cwetanski	54
Rita De Masi	3
Biagio Di Micco	33
Tatyana Donskova	2
Eric Dumonteil	55
Henning Flaecher	13
Robert Fleischer	60
Dmitry Fursaev	58
Bruce Gallop	17
Olivier Gaumer	56
Chiara Genta	11
Nils Gollub	68
Lukasz Goscilo	25
Andranik Hakobyan	42
Josef Juran	29
Vladimir Kadyshevsky	67
Gia Khoriauli	48
Vigen Kojoyan	4
Venelin Kozhuharov	10
Marcin Kucharczyk	75
Marianna Kuznetsova	47
Anne-Catherine Le Bihan	6
Marie Legendre	20
Egil Lillestøl	16
Andrew Lowe	78
Katharina Mair	30
Ewa Markiewicz	7
Adam Matyja	59
Danielle Métral	24
Niels Meyer	51
Alessandro Montanari	12
Carsten Noeding	71
Emily Nurse	64
Alexander Olchevsky	39
Bakur Parsamyan	49
Ralf Patrick Bernhard	74
Charles Pattison	35
Ciro Pistillo	32
Alexei Pivovarov	41
Tilman Plehn	73
Michael Pluemacher	66
Michael Poettgens	14
George Pogosyan	40
Marco Poli Lener	52
Robindra Prabhu	15
Davtyan Razmik	19
Melisa Rossi	70
Luigi Salvatore Esposito	21
Luca Scotto Lavina	34
Radomir Smida	23
Ivan Sotsky	1
Krzysztof Syryczynski	50
Oleg Teryaev	18
Silvano Tosi	45
Piotr Traczyk	26
Roberto Versaci	27
Sébastien Viret	46
Tuan Vu Anh	9
Ronald Weber	31
Armen Yeranyan	37
Viatcheslav Zholobov	76
Stéphanie Zimmermann	28

CONTENTS

Preface
E. Lillestøl v

Photograph of participants vi

Field theory and Standard Model
I.J.R. Aitchison 1

Galactic and solar cosmic rays
A. Chilingarian 63

Flavour physics and CP violation
R. Fleischer 81

Beyond the Standard Model: extra dimensions and supersymmetry
G. Gabadadze 151

Quantum chromodynamics and hadrons: an elementary introduction
A. Khodjamirian 173

On the thermodynamics of inelastic hadron processes
I. Manjavidze and A. Sissakian 223

Astroparticle physics
I. Tkachev 253

Organizing Committee 307
Lecturers 307
Discussion Leaders 307
Other Attendees 307
Students and Listeners 308
List of Posters 309

FIELD THEORY AND STANDARD MODEL

I.J.R. Aitchison
Department of Physics, University of Oxford, UK

Abstract
This is a course of six lectures given at the 2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, 24 August–6 September 2003. They aim to provide a compact introduction to quantum field theory (in the 'canonical' formalism) and the Standard Model, focusing on field quantization and the canonical route to the Feynman rules; Abelian symmetries and QED; one-loop renormalization of QED; non-Abelian symmetries; spontaneously broken symmetries; and the electroweak theory.

1 OUTLINE OF THE COURSE

Section 2 (Lecture 1) Canonical quantization of free spin-0 (scalar) field. Interacting scalar fields. The Dyson-Wick expansion of the S-matrix. Propagators. Tree graphs. The Yukawa potential.
Section 3 (Lecture 2) Complex scalar field. Global U(1) phase invariance. Number conservation laws. Fermions. Local U(1) phase invariance and the electromagnetic interaction. The Maxwell field. Elements of QED.
Section 4 (Lecture 3) One-loop graphs in QED: renormalization, and running coupling constant.
Section 5 (Lecture 4) Non-Abelian symmetries, global and local. Local SU(2) symmetry. Gauge field self-interactions. Local SU(3) symmetry. QCD.
Section 6 (Lecture 5) Spontaneous symmetry breaking, global and local. Chiral symmetry breaking. The Abelian Higgs model. Spontaneously broken SU(2)×U(1).
Section 7 (Lecture 6) Introduction to the electroweak theory. The Higgs sector. One loop effects.

2 SCALAR FIELDS: TO TREE GRAPHS

A more leisurely treatment of the material in this section is given in Chapters 5 and 6 of Volume 1 of the new (third) edition of Aitchison and Hey [1].

2.1 The classical field as an assembly of non-interacting oscillators

Consider a familiar problem, that of a string stretched between points $x = 0$ and $x = L$. The transverse displacement y of the string at position x and time t, $y(x,t)$, satisfies the wave equation

$$\frac{\partial^2 y(x,t)}{c^2 \partial t^2} = \frac{\partial^2 y(x,t)}{\partial x^2} \tag{1}$$

for small displacements. Here $y(x,t)$ is a scalar field: 'scalar' because it has only one component, and 'field' because it varies continuously with x and t. The fundamental method of solving equations like (1) is first to find particular solutions called *modes*, and then to use the fact that (1) is *linear* to write the *general solution as a linear superposition of modes*. Here, the modes must satisfy the boundary conditions $y(0,t) = y(L,t) = 0$, so we try

$$y(x,t) = X_r(t) \sin(\frac{r\pi x}{L}) \tag{2}$$

for $r = 1, 2...$, which expresses the fact that any number of half-wavelengths must fit into the interval $(0, L)$. Substituting (2) into (1) we find

$$X_r(t) = -\omega_r^2 X_r(t) \tag{3}$$

where

$$\omega_r^2 = c^2 r^2 \pi^2 / L^2. \tag{4}$$

Thus each *mode amplitude* $X_r(t)$ executes simple harmonic motion with frequency $\omega_r = (cr\pi/L)$: it acts like the 'coordinate' of an oscillator! The general solution of (1) is then

$$y(x,t) = \sum_{r=1}^{\infty} X_r(t) \sin(\frac{r\pi\ x}{L}); \tag{5}$$

in short, a Fourier series.

Now let us consider the total energy of the vibrating string, which is given by the integral

$$E = \int_0^L [\frac{1}{2}\rho \left(\frac{\partial\, y}{\partial t}\right)^2 + \frac{1}{2}\rho c^2 \left(\frac{\partial\, y}{\partial\, x}\right)^2] dx, \tag{6}$$

where the first term is the kinetic energy 'T' (ρ is the mass per unit length) and the second is the potential energy 'V'. When (5) is placed in (6) and the integral over x done, a remarkable result is obtained (problem P1.1):

$$E = \frac{L}{2} \sum_{r=1}^{\infty} [\frac{1}{2}\ \rho \dot{X}_r^2 + \frac{1}{2}\ \rho\ \omega_r^2\ X_r^2]. \tag{7}$$

Equation (7) has a strikingly simple physical interpretation: the energy of the string is equal to the sum of the energies of individual 'mode oscillators' (recall the energy of one simple harmonic oscillator (SHO) is $\frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2$ so here $X_r \leftrightarrow x, \frac{L\rho}{2} \leftrightarrow m, \omega_r \leftrightarrow \omega$). For a general motion of the strings, all the oscillators (modes) will be present. Because the total energy is the *sum* of the individual mode energies, the *modes do not interact with each other*. So, from the point of view of the energy, at least, *the field is equivalent to an assembly of non-interacting oscillators*.

2.2 Quantization

Let us write $M = \frac{L\rho}{2}$ so that (7) becomes

$$E = \sum_r \frac{1}{2} M \dot{X}_r^2 + \frac{1}{2} M w_r^2 X_r^2. \tag{8}$$

The essential idea is to treat the mode amplitudes 'X_r' as 'quantum coordinate-like variables'. The associated 'momentum-like variables' will be $P_r = M\dot{X}_r$. The energy (8) (which of course in classical physics is a number) becomes now an *operator*, namely the Hamiltonian operator

$$\hat{H} = \sum_{r=1}^{\infty} \frac{\hat{P}_r^2}{2\ M} + \frac{1}{2} M \omega_r^2\ \hat{X}_r^2. \tag{9}$$

We know all about the energy levels and states of a *single* quantum oscillator; the fact that we have here arbitrarily many oscillators does not worry us as they are not interacting with each other, so they can be treated quite independently. For a single oscillator of frequency ω, the energy levels are $E_n = (n + \frac{1}{2})\ \hbar\omega$, and the wavefunctions are well-known, in all quantum mechanics textbooks. For our purposes, we prefer the 'operator approach' in terms of $\hat{a}$'s and $\hat{a}^\dagger$'s to the wavefunction one. The essentials are gone through in problem P1.2.

For our vibrating string, then, we simply have

$$\hat{H} = \frac{1}{2} \sum_{r=1}^{\infty} (\hat{a}_r^\dagger\ \hat{a}_r\ + \hat{a}_r\ \hat{a}_r^\dagger)\ \hbar\omega_r, \quad \text{with} \quad [\hat{a}_r, \hat{a}_s^\dagger] = \delta_{rs}. \tag{10}$$

The eigenstates of $\hat{H}$ are *products* of the single oscillator states $|n_1\rangle|n_2\rangle|n_3\rangle\ldots\ldots$ where $|n_1\rangle$ is the state of the oscillator with frequency ω_1, which has energy $(n_1+\frac{1}{2})\hbar\omega_1$, etc. We can write this more briefly as $|n_1, n_2, ...\rangle$, which has energy $\sum_r (n_r+\frac{1}{2})\hbar\omega_r$. The ground state $|0\rangle$ has *all* n_r's $= 0$, and hence an energy (the 'zero point energy') equal to $\sum_r \frac{1}{2}\hbar\omega_r$.

Thus the energy eigenstates of the quantized field $\hat{y}(x,t)$ are characterized by saying *how many quanta of each frequency* are present; in the ground state there are *no* quanta of excitation present. Such *vibrational quanta* are called 'phonons' in condensed matter physics. *Our 'particles' are similar quanta of excitation of fields*. The state with no excitation quanta is a (too simple!) model of the *vacuum*.

2.3 Free massive real scalar field

We will from now on put $\hbar = c = 1$. The 'classical' field satisfies the Klein–Gordon equation

$$(\Box + m^2)\phi = (\frac{\partial^2\phi}{\partial\, t^2} - \boldsymbol{\nabla}^2\phi + m^2)\phi = 0, \quad \text{where } \Box = \partial_\mu\partial^\mu = \partial_t^2 - \boldsymbol{\nabla}^2 \tag{11}$$

which is the wave equation for a free massive spin-0 (scalar) field. We now consider the field to be in 'infinite space' so Fourier series $\rightarrow$ Fourier integrals and our *modes* have the form

$$\phi(\boldsymbol{x}, t) = X_{\boldsymbol{k}}(t)\exp[i\boldsymbol{k}\cdot\boldsymbol{x}]. \tag{12}$$

Plugging this into the Klein–Gordon (KG) equation gives

$$\ddot{X}_{\boldsymbol{k}} = -(m^2 + \boldsymbol{k}^2)X_{\boldsymbol{k}} \tag{13}$$

which again shows that our mode amplitude acts like an SHO, this time with frequency

$$\omega_k = \pm(m^2 + \boldsymbol{k}^2)^{\frac{1}{2}}. \tag{14}$$

The total energy in the field is the obvious generalization of the energy of the string:

$$E = \frac{1}{2}\int d^3\boldsymbol{x}[\dot{\phi}^2 + (\boldsymbol{\nabla}\phi)^2 + m^2\phi^2]. \tag{15}$$

Once again, this can be written as a 'sum' (in this case, an integral over the Fourier variable $\boldsymbol{k}$) of independent energies for each mode oscillator. So, when *quantized*, we get the Hamiltonian [compare Eq. (10)]

$$\hat{H}_{KG} = \frac{1}{2}\int \frac{d^3\boldsymbol{k}}{(2\pi)^3}k_0\{\hat{a}^\dagger(\boldsymbol{k})\hat{a}(\boldsymbol{k}) + \hat{a}(\boldsymbol{k})\hat{a}^\dagger(\boldsymbol{k})\} \tag{16}$$

where $k_0 = +(m^2 + \boldsymbol{k}^2)^{\frac{1}{2}}$, and where the mode creation and annihilation operators satisfy

$$[\hat{a}(\boldsymbol{k}), \hat{a}^\dagger(\boldsymbol{k}')] = (2\pi)^3\delta(\boldsymbol{k} - \boldsymbol{k}'), \tag{17}$$

all other commutators vanishing: $[\hat{a}^\dagger(\boldsymbol{k}), \hat{a}^\dagger(\boldsymbol{k}')] = [\hat{a}(\boldsymbol{k}), \hat{a}(\boldsymbol{k}')] = 0$.

Problem P1.3 shows that the state $|p\rangle \propto \hat{a}^\dagger(\boldsymbol{p})|0\rangle$ is an eigenstate of $\hat{H}$ with eigenvalue $\sqrt{m^2 + \boldsymbol{p}^2}$, the expected energy for a particle of mass m and momentum $\boldsymbol{p}$ (note $\hbar = c = 1$). We actually choose the *particular normalization*

$$|p\rangle = \sqrt{2p_0}\hat{a}^\dagger(\boldsymbol{p})|0\rangle. \tag{18}$$

The general (quantized) solution to the KG field equation is then

$$\hat{\phi}(x) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3(2k_0)^{\frac{1}{2}}}\{\hat{a}(\boldsymbol{k})\exp[-ik\cdot x] + \hat{a}^\dagger(\boldsymbol{k})\exp[ik\cdot x]\} = \hat{\phi}^\dagger(x) \tag{19}$$

for a 'real' field, and where $k \cdot x = k_0 x_0 - \boldsymbol{k} \cdot \boldsymbol{x}$ and the $(2k_0)^{-\frac{1}{2}}$ is a conventional normalization factor.

Problem P1.4 shows that

$$\langle 0|\hat{\phi}(x)|p\rangle = e^{-ip\cdot x}. \tag{20}$$

In ordinary quantum mechanics the RHS of this equation would be written as $\langle x|p\rangle$, the x-space wavefunction for a state $|p\rangle$ of definite 4-momentum p (which is of course a 4-D plane wave). We can then reasonably assert that the operator

$$\int \frac{d^3\boldsymbol{k}}{(2\pi)^3(2k_0)^{\frac{1}{2}}}\hat{a}^\dagger(\boldsymbol{k})e^{ik\cdot x} \equiv \hat{\phi}^{(-)}(x) \tag{21}$$

creates a quantum at x: $\hat{\phi}^{(-)}(x)|0\rangle = |x\rangle$. (Note that the *other* part of $\hat{\phi}(x)$ gives 0 when acting on $|0\rangle$).

The commutation relations (17) imply that

$$|p_1, p_2\rangle = |p_2, p_1\rangle \tag{22}$$

so our particles are bosons!

2.4 Interactions

In the case of the freely vibrating string, or the free scalar field, the energy is the *sum* of individual mode energies—the modes do not interact. But our particles are precisely mode quanta, and we want them to interact, of course. So we must complicate our simple expressions for field energies in some way. The crucial feature of (6) and (15) which leads to the '$\sum$ mode energies' result is that they are *quadratic* in the fields and their derivatives. *Interactions* will generally be represented by expressions which are cubic or quartic in the fields. Correspondingly, quadratic or cubic expressions will appear in the equations of motion. (Compare the SHO: the 'free' SHO energy is $\frac{1}{2}mx^2 + \frac{1}{2}m\omega^2x^2$ with equation of motion $m\ddot{x} = -m\omega^2 x$; if it is perturbed by adding a cubic potential energy λx^3, this produces a force $-\frac{dV}{dx} = -3\lambda x^2$.) In the case of lattice vibrations, such 'anharmonic terms' cause the *phonons to interact*—it is the same with our particles. We will introduce an interaction term $\hat{H}'$ in the Hamiltonian:

$$\hat{H}' = \int d^3\boldsymbol{x}\hat{\mathcal{H}}'(\hat{\phi}), \tag{23}$$

for example

$$\hat{\mathcal{H}}' = \lambda(\hat{\phi}(x))^3. \tag{24}$$

We treat $\hat{H}'$ as a perturbation on $\hat{H}_{KG}$.

2.5 Covariant perturbation theory: the Dyson–Wick expansion of the $\hat{S}$ operator, Feynman rules

There is a very compact and powerful formalism for doing relativistic perturbation theory, which we are not going to go through the details of here—just quote the essential results. Transitions are described by means of matrix elements (between free-particle states $|i\rangle$ and $|f\rangle$) of the $\hat{S}$ operator, $\langle f|\hat{S}|i\rangle$, where $\hat{S}$ has the expansion in powers of $\hat{\mathcal{H}}'$:

$$\hat{S} = 1 - i\int d^4x\hat{\mathcal{H}}'(\hat{\phi}(x')) + \frac{1}{2}\int\int d^4x_1 d^4x_2 T(-i\hat{\mathcal{H}}'(x_1). - i\hat{\mathcal{H}}'(x_2)) + \ldots \tag{25}$$

where 'T' is the time-ordering operation

$$\begin{aligned} T(\hat{\phi}(x_1)\hat{\phi}(x_2)) &= \hat{\phi}(x_1)\hat{\phi}(x_2) \text{ for } t_1 > t_2 \\ &= \hat{\phi}(x_2)\hat{\phi}(x_1) \text{ for } t_1 < t_2 \end{aligned} \tag{26}$$

i.e. 'earlier on the right'.

Discussion Point: This is supposed to be covariant (relativistically invariant) perturbation theory. But the 'T' symbol seems to be singling out 'time' in some way, and does not look '4-D symmetric'. Should we be worried?

Example: 'ABC' theory

To have a little more variety than the single $\hat{\phi}$ field, let us imagine a world with three real scalar fields $\hat{\phi}_A(\text{mass } m_A), \hat{\phi}_B(m_B)$ and $\hat{\phi}_C(m_C)$ with an interaction $g\hat{\phi}_A(x)\hat{\phi}_B(x)\hat{\phi}_C(x)$. This interaction creates or annihilates one each of an A,B or C particle—for example C $\rightarrow$ A + B. Suppose $m_C > m_A + m_B$. Then C will be able to decay to A + B. The matrix element for this will be, to lowest order,

$$\int d^4x \langle p_A, p_B| - ig\hat{\phi}_A(x)\hat{\phi}_B(x)\hat{\phi}_C(x)|p_C\rangle. \tag{27}$$

Problem P1.5 shows that this matrix element is equal to $-ig(2\pi)^4\delta^4(p_C - p_A - p_B)$. (Note: creation and annihilation operators for the *different* fields commute with each other.) So we have our first 'Feynman rule'!

(i) $-ig$ for an 'A-B-C' vertex together with an overall factor of $(2\pi)^4\delta(p_{\text{initial}} - p_{\text{final}})$.

Now consider A B $\rightarrow$ A B scattering. The lowest order in perturbation theory at which this process can proceed is *second*, via the matrix element

$$\frac{1}{2}\iint d^4x_1 d^4x_2 \langle p'_A, p'_B|T\{(-ig\hat{\phi}_A(x_1)\hat{\phi}_B(x_1)\hat{\phi}_C(x_1))(-ig\hat{\phi}_A(x_2)\hat{\phi}_B(x_2)\hat{\phi}_C(x_2))\}|p_A, p_B\rangle. \tag{28}$$

Suddenly we have a complicated expression on our hands! Remembering (18), we see that (28) is essentially

$$(16E_AE_BE'_AE'_B)^{\frac{1}{2}}\langle 0|\hat{a}_A(p'_A)\hat{a}_B(p'_B)T\{\hat{\phi}_A(x_1)\hat{\phi}_B(x_1)\hat{\phi}_C(x_1)\hat{\phi}_A(x_2)\hat{\phi}_B(x_2)\hat{\phi}_C(x_2)\}\hat{a}^\dagger_A(p_A)\hat{a}^\dagger_B(p_B)|0\rangle \tag{29}$$

which is the vacuum expectation value (vev) of 10 operators. Remarkably, it can be shown (*Wick's theorem*) that such vev's can be written as a sum of products of all possible choices of *pairwise vev's* (time-ordered vev's, in general). *One* such term is

$$\begin{aligned}\iint d^4x_1 d^4x_2 &\langle 0|\hat{a}_A(p'_A)\hat{\phi}_A(x_1)|0\rangle\langle 0|\hat{\phi}_A(x_2)\hat{a}^\dagger_A(p_A)|0\rangle\langle 0|\hat{a}_B(p'_B)\hat{\phi}_B(x_2)|0\rangle\times \\ &\times\langle 0|\hat{\phi}_B(x_1)\hat{a}^\dagger_B(p_B)|0\rangle\langle 0|T(\hat{\phi}_C(x_1)\hat{\phi}_C(x_2))|0\rangle \times (16E_AE_BE'_AE'_B)^{\frac{1}{2}}.\end{aligned} \tag{30}$$

Problem 1.4 shows us that the terms with one field and one $\hat{a}$ or $\hat{a}^\dagger$ give just plane waves: two ingoing ones and two outgoing ones, yielding $\exp i\{p'_A \cdot x_1 - p_A \cdot x_2 + p'_B \cdot x_2 - p_B \cdot x_1\}$. The *interesting* bit is the remaining *vev of the time-ordered product of two $\hat{\phi}$ fields*, which is the *Feynman propagator* in coordinate space. The physical interpretation of the *two* terms in $\langle 0|T(\hat{\phi}_C(x_1)\hat{\phi}_C(x_2))|0\rangle$, one for $t_1 > t_2$ and one for $t_1 < t_2$ is as follows: A C-quantum is being produced at x_1 and destroyed by x_2, *or* the other way round (*Exercise:* explain why, with the aid of the mode expansion for $\hat{\phi}_C(x)$). So including the incoming and outgoing plane waves we have the physical processes shown in Fig. 1, and we have to integrate the whole expression in (30) *over all x_1 and x_2*. The result is the Feynman rule in *momentum space* for the scalar propagator (see textbooks):

(ii) a factor $i/[(\text{4-momentum carried by the propagating particle})^2 - (\text{its mass})^2]$

So for the C-exchange process we have the diagram of Fig. 2, corresponding to the Feynman amplitude $-ig/(q^2 - m_C^2)$ where $q = p_A - p'_B$. In addition, there is the overall factor $(2\pi)^4\delta^4(p_A + p_B - p'_A - p'_B)$.

Points to note:

1 A and B are interacting by 'exchanging a C'.

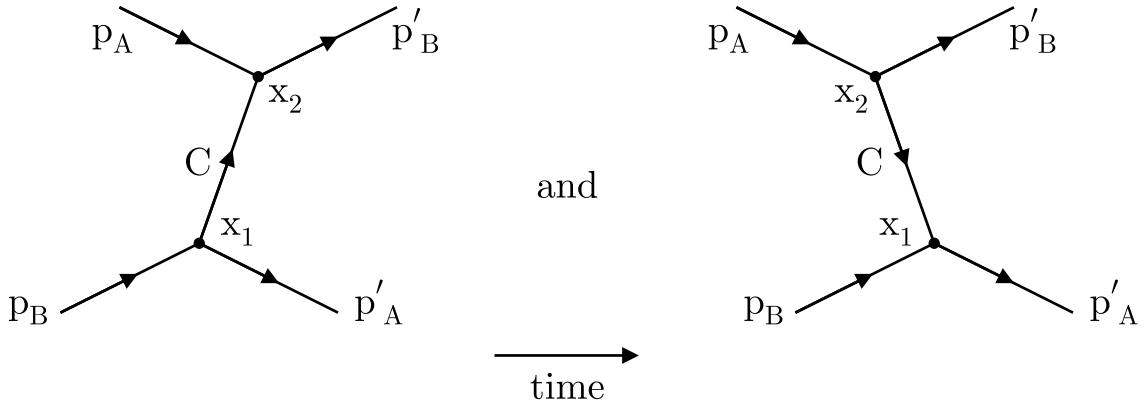


Fig. 1: The two physical processes included in the single Feynman C propagator.

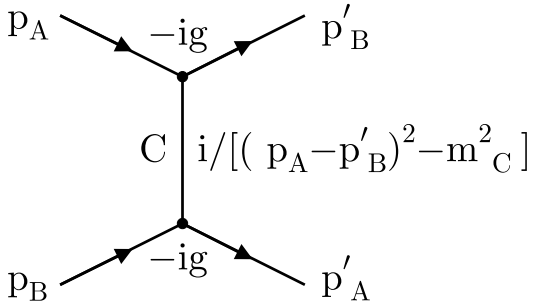


Fig. 2: One-C exchange process in A + B → A + B.

2 But the (4-momentum)^2 carried by the exchanged C is *not* equal to m_C^2—it is 'off mass shell'.

3 *Both* time orderings are included in this *one* momentum space amplitude.

4 Suppose we evaluate the amplitude in the c.m. frame: $p_A = (E_A, \boldsymbol{p}), p_B = (E_B, -\boldsymbol{p}), p'_A = (E_A, -\boldsymbol{p}'), p'_B = (E_B, +\boldsymbol{p}'), |\boldsymbol{p}| = |\boldsymbol{p}'|$. Then $(p_A - p'_B)^2 = (E_A - E_B)^2 - (\boldsymbol{p} - \boldsymbol{p}')^2$. Now consider the static or non-relativistic limit $(E_A - E_B)^2 \ll (\boldsymbol{p} - \boldsymbol{p}')^2$. Our amplitude is now essentially

$$\sim \frac{1}{(\boldsymbol{p} - \boldsymbol{p}')^2 + m_C^2}. \tag{31}$$

We can interpret this in terms of a *potential* associated with the A-B interaction. According to the Born approximation in scattering theory, the amplitude to go from $\boldsymbol{p}$ to $\boldsymbol{p}'$ in the potential $V(\boldsymbol{r})$ is

$$\sim \int \exp\{-i\boldsymbol{p}' \cdot \boldsymbol{r}\} V(\boldsymbol{r}) \exp\{i\boldsymbol{p} \cdot \boldsymbol{r}\} d^3\boldsymbol{r} = \int \exp\{i(\boldsymbol{p} - \boldsymbol{p}') \cdot \boldsymbol{r}\} V(\boldsymbol{r}) d^3\boldsymbol{r} \tag{32}$$

which is some function of $(\boldsymbol{p} - \boldsymbol{p}')^2$. Question: what is $V(\boldsymbol{r})$ such that this function is equal to $[(\boldsymbol{p} - \boldsymbol{p}')^2 + m_C^2]^{-1}$? Answer: $V(\boldsymbol{r}) \propto \exp\{-m_C|\boldsymbol{r}|\}/|\boldsymbol{r}|$, the *Yukawa potential*, of range $1/m_C$; see problem P1.6.

5 Good exercise: think about some of the *other* terms in the Wick expansion of (28)!

Problems for Lecture 1

P1.1 A string is stretched between two points $x = 0$ and $x = L$. The transverse displacement of the string at the point x at time t is $y(x, t)$ where

$$\frac{\partial^2 y}{c^2 \partial t^2} = \frac{\partial^2 y}{\partial x^2}.$$

The general solution can be written as a superposition

$$y(x,t) = \sum_{r=1}^{\infty} X_r(t) \sin \frac{r\pi x}{L}.$$

The total energy of the vibrating string is

$$E = \int_0^L [\frac{1}{2}\rho \left(\frac{\partial y}{\partial t}\right)^2 + \frac{1}{2}\rho c^2 \left(\frac{\partial y}{\partial x}\right)^2] dx$$

where ρ is the mass per unit length. Show that

$$E = \frac{L}{2} \sum_r [\frac{1}{2}\rho \dot{X}_r^2 + \frac{1}{2}\rho \omega_r^2 X_r^2]$$

where $\omega_r = cr\pi/L$. [Hint: write the term $(\frac{\partial y}{\partial t})^2$, for example, as a product of *two independent summations* $(\sum_r \ldots)(\sum_s \ldots)$ and explain why there are no 'cross terms' of the form $X_r X_s \;\; r \neq s$ in the answer.]

P1.2 A one-dimensional harmonic oscillator has the Hamiltonian (energy operator) $\hat{H} = \hat{p}^2/2m + \frac{1}{2}\omega^2 \hat{x}^2$ where $[\hat{x}, \hat{p}] = i$ (units $\hbar = 1$). Define the operators $\hat{a}, \hat{a}^\dagger$ by

$$\hat{a} = \sqrt{\frac{m\omega}{2}} \left(\hat{x} + \frac{i\hat{p}}{m\omega}\right), \quad \hat{a}^\dagger = \sqrt{\frac{m\omega}{2}} \left(\hat{x} - \frac{i\hat{p}}{m\omega}\right).$$

(i) Show that $[\hat{a}, \hat{a}^\dagger] = 1$. (ii) Show that $\hat{H}$ can be written as $\frac{1}{2}\hbar\omega(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})$ or as $\hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2})$. Deduce that $[\hat{a}^\dagger, \hat{H}] = -\omega\hat{a}^\dagger$ and hence show that if $\hat{H}|n\rangle = E_n|n\rangle$ then $\hat{H}\hat{a}^\dagger|n\rangle = (E_n + \omega)|n\rangle$, so that $\hat{a}^\dagger|n\rangle \propto |n+1\rangle$. State and prove a similar result for $\hat{a}|\, n\rangle$. (iii) Explain why there must be a state $|0\rangle$ such that $\hat{a}|0\rangle = 0$. What is the energy eigenvalue of $|0\rangle$? Deduce the energy spectrum of the oscillator.

P1.3 $\hat{H}_{KG}$ is defined by

$$\hat{H}_{KG} = \frac{1}{2} \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} k_0 \{\hat{a}^\dagger(\boldsymbol{k})\hat{a}(\boldsymbol{k}) + \hat{a}(\boldsymbol{k})\hat{a}^\dagger(\boldsymbol{k})\}$$

where $k_0 = +\sqrt{m^2 + \boldsymbol{k}^2}$, with

$$[\hat{a}(\boldsymbol{k}), \hat{a}^\dagger(\boldsymbol{k'})] = (2\pi)^3 \delta(\boldsymbol{k} - \boldsymbol{k'})$$

all other commutators vanishing. Show that

$$\hat{H}_{KG}\hat{a}^\dagger(\boldsymbol{p})|0\rangle = p_0 \hat{a}^\dagger(\boldsymbol{p})|0\rangle$$

where $p_0 = +\sqrt{m^2 + \boldsymbol{p}^2}$ and $\hat{a}(\boldsymbol{k})|0\rangle = 0$ for all $\boldsymbol{k}$.

P1.4 The field $\hat{\phi}(x)$ has the mode expansion

$$\hat{\phi}(x) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3 (2k_0)^{\frac{1}{2}}} \{\hat{a}(\boldsymbol{k}) \exp[-ik \cdot x] + \hat{a}^\dagger(\boldsymbol{k}) \exp[ik \cdot x]\}$$

where $k \cdot x = k_0 x_0 - \boldsymbol{k} \cdot \boldsymbol{x}$. Show that

$$\langle 0|\hat{\phi}(x)|p\rangle = e^{-ip\cdot x}$$

where

$$|p\rangle = \sqrt{2p_0}\hat{a}^\dagger(\boldsymbol{p})|0\rangle.$$

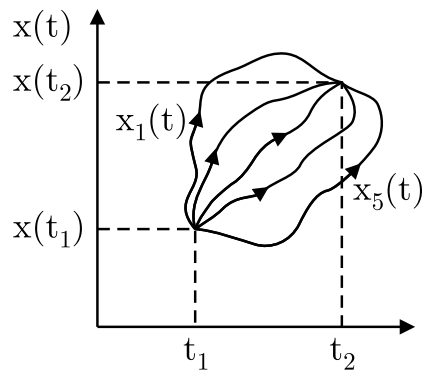


Fig. 3: Possible space–time trajectories between the fixed points $x(t_1)$ and $x(t_2)$.

P1.5 $\hat{\phi}_A, \hat{\phi}_B,$ and $\hat{\phi}_C$ are three distinct scalar fields. Evaluate

$$\int d^4x \langle p_A, p_B| - ig\hat{\phi}_A(x)\hat{\phi}_B(x)\hat{\phi}_C(x)|p_C\rangle.$$

P1.6 Evaluate the Fourier transform

$$\int d^3\boldsymbol{r} \exp\{i\boldsymbol{q}\cdot\boldsymbol{r}\}\frac{\exp\{-r/a\}}{r}$$

of a Yukawa potential by following these steps: change $d^3\boldsymbol{r}$ to polar coordinates '$r^2 dr \sin\theta d\theta d\phi$' with the polar axis chosen along the direction of $\boldsymbol{q}$. So $\exp\{i\boldsymbol{q}\cdot\boldsymbol{r}\} = \exp\{i|\boldsymbol{q}|r\cos\theta\}$. Do the integral over θ. Then do the integral over r (the ϕ integral just gives 2π). [Answer: $4\pi/(\boldsymbol{q}^2 + a^{-2})$.]

3 LAGRANGIANS, COMPLEX SCALAR FIELDS, DIRAC AND MAXWELL FIELDS

See Chapter 7 of Ref. [1].

We have managed to get this far without mentioning the word 'Lagrangian', but now we are going to have to start using this language, which is particularly well suited to the discussion of *symmetries*, and these are of fundamental importance in the Standard Model (SM).

3.1 Lagrangians

This is essentially a formulation of dynamics which is different from (but in the classical case equivalent to) Newton's. The basic quantity here is the Lagrangian function, which in most cases has the form '$L = T - V$' (instead of the energy which is '$E = T + V$'). For a classical particle with coordinate $x(t)$, L is just $L[x(t), \dot{x}(t)] = \frac{1}{2}m\dot{x}(t)^2 - V(x(t))$. The 'path' $x(t)$ the particle takes is determined by the *principle* that the *action integral* S given by

$$S = \int_{t_1}^{t_2} L[x(t), \dot{x}(t)] = \int_{t_1}^{t_2} [\frac{1}{2}m\dot{x}(t)^2 - V(x(t))]dt \tag{33}$$

is a minimum as all paths $x(t)$ are searched over, subject to $x(t_1)$ and $x(t_2)$ being fixed (see Fig. 3). Problem P2.1 provides a simple example.

Although the action principle seems very different from the *differential equations* of Newton's laws, we can connect them by using a bit of calculus. The actual path must be determined from the condition that small changes away from it make no change in S, to first order (i.e. S is at a minimum).

So consider an arbitrary change $x(t) \to x(t) + \delta x(t)$, which also implies $\dot{x}(t) \to \dot{x}(t) + \frac{d}{dt}\delta x(t)$. So then $\dot{x}^2 \to \dot{x}^2 + 2\dot{x}\frac{d}{dt}\delta x$ to first order, and $V(x) \to V(x) + \frac{dV}{dx}\delta x$, giving

$$\delta S = \int_{t_1}^{t_2} [m\dot{x}\frac{d}{dt}(\delta x) - \frac{dV}{dx}\delta x]dt. \tag{34}$$

Now do a partial integration in the first term to get

$$\delta S = \int_{t_1}^{t_2} -[\frac{d}{dt}(m\dot{x}) + \frac{dV}{dx}]\delta x(t)dt, \tag{35}$$

assuming that δx vanishes at the end points (all paths start and finish at the same points). Now it is important to realise that '$\delta x(t)$' here is an *arbitrary* (if 'small') *function of* t. But this change in S, δS, must be zero, by our principle. The only way the integral in (35) can be zero for *arbitrary* $\delta x(t)$ is if the quantity inside the square brackets vanishes, i.e.

$$\frac{d}{dt}(m\dot{x}) = -\frac{dV}{dx} \tag{36}$$

which is exactly Newton's law of motion!

In quantum mechanics, the action approach can also be used, as stated by Dirac and developed by Feynman. There, the amplitude to go from $x(t_1)$ to $x(t_2)$ is proportional to

$$\sum_{\text{all paths } x(t)} \exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} L(x(t),\dot{x}(t))dt\right) = \sum_{\text{paths}} \exp iS/\hbar. \tag{37}$$

The qualitative idea here is that if the integral is an essentially classical quantity, then its value will be a very large number of $\hbar$'s, so the phase factor will oscillate wildly as the x's change, and everything will cancel out *except* for trajectories such that the action is stationary to small variations around them, since for these ones the phases will 'add up' coherently; hence we get back to the classical action principle in that case.

The action approach can also be used for fields, both classical and quantum; for the latter, see Peter Hasenfratz's lectures. In this course we shall not use it for dynamics (i.e. for deriving the Feynman rules), but we shall use the Lagrangian language, because it is a very powerful one for discussing symmetries, and because it is quite simply the *lingua franca* of particle physics (at least insofar as the Standard Model is concerned). Before moving to that, we note that the general formulation of (36) is (problem P2.2)

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0. \tag{38}$$

For fields, we shall have to introduce a Lagrangian *density* $\mathcal{L}$ such that (in one space dimension)

$$S = \int\int dtdx\mathcal{L}[\phi(x), \dot{\phi} = \frac{\partial\phi(x)}{\partial t}, \frac{\partial\phi(x)}{\partial x}]; \tag{39}$$

$\dot{\phi} = \frac{\partial\phi}{\partial t}$ is like $\dot{x}$ in (33), and $\frac{\partial\phi}{\partial x}$ is new, but analogous. Again, the field equation for $\phi(x)$ will be determined from the condition that $\delta S = 0$ under $\phi \to \phi + \delta\phi$, $\dot{\phi} \to \dot{\phi} + \delta\dot{\phi}$, $\frac{\partial\phi}{\partial x} \to \frac{\partial\phi}{\partial x} + \delta\left(\frac{\partial\phi}{\partial x}\right)$:

$$0 = \delta S = \int dt \int \left[\frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial\phi/\partial x)}\delta\left(\frac{\partial\phi}{\partial x}\right) + \frac{\partial\mathcal{L}}{\partial\dot{\phi}}\delta\dot{\phi}\right] dx. \tag{40}$$

(Compare (34).) But $\delta\left(\frac{\partial\phi}{\partial x}\right) = \frac{\partial}{\partial x}\delta\phi$, and similarly for the $\dot{\phi}$ term, so that the second and third terms in (40) can both be integrated by parts, as in (35). As in that case, the variations vanish at the end-points, and since $\delta\phi$ is arbitrary, we deduce the *Euler–Lagrange equation of motion*

$$\frac{\partial\mathcal{L}}{\partial\phi} - \frac{\partial}{\partial x}\left(\frac{\partial\mathcal{L}}{\partial(\partial\phi/\partial x)}\right) - \frac{\partial}{\partial t}\left(\frac{\partial\mathcal{L}}{\partial\dot{\phi}}\right) = 0. \tag{41}$$

Example. $\mathcal{L} = \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\frac{\partial\phi}{\partial x})^2 - \frac{1}{2}m^2\phi^2$. The E–L equation is $\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} + m^2\phi = 0$, the KG equation.

This all generalizes to 4-D via

$$\mathcal{L}_{\rm KG} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \tag{42}$$

Here $\partial^\mu = \frac{\partial}{\partial x_\mu}, \partial_\mu = \frac{\partial}{\partial x^\mu}, x^\mu = (x^0, \boldsymbol{x}), x_\mu = (x^0, -\boldsymbol{x}), \partial^\mu\partial_\mu = \frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x_\mu} = \frac{\partial}{\partial x^0}\frac{\partial}{\partial x^0} + \frac{\partial}{\partial \boldsymbol{x}}\cdot(-\frac{\partial}{\partial \boldsymbol{x}}) = \frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2$.

And this generalizes to *quantum fields* by putting hats on!

3.2 The complex scalar field

In Section 2 we considered a 'real' scalar field for which $\hat{\phi}^\dagger = \hat{\phi}$. The next most complicated thing is a complex scalar field for which $\hat{\phi}^\dagger$ is different from $\hat{\phi}$. So here our mode expansion will have the form

$$\hat{\phi} = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3\sqrt{2\omega}}[\hat{a}(k)e^{-ik\cdot x} + \hat{b}^\dagger(k)e^{ik\cdot x}] \ . \tag{43}$$

The physical interpretation of this is that '$\hat{a}$' will destroy a *particle* (quantum) of the field, while '$\hat{b}^\dagger$' will create an *antiparticle*. This is because states $\hat{a}^\dagger|0\rangle$ and $\hat{b}^\dagger|0\rangle$ are distinguished by having opposite signs of a certain *conserved quantum number*. Now conservation laws have to do with symmetries: what symmetry is at work here? The answer is that it is a symmetry under

$$\hat{\phi} \to e^{-i\alpha}\hat{\phi} \tag{44}$$

i.e. a simple phase transformation. Any $\hat{\mathcal{L}}(\hat{\phi})$ which is a function of $\hat{\phi}^\dagger\hat{\phi}$ and $\partial_\mu\hat{\phi}^\dagger\partial^\mu\hat{\phi}$ only will be invariant (symmetric) under (44); for instance the Lagrangian for the free complex KG field

$$\hat{\mathcal{L}} = \partial_\mu\hat{\phi}^\dagger\partial^\mu\hat{\phi} - m^2\hat{\phi}^\dagger\hat{\phi} \tag{45}$$

is invariant under (44).

The symmetry (44) is called a *continuous* symmetry because the phase angle α can be anything (compare 'parity', where the transformation is $\boldsymbol{x} \to -\boldsymbol{x}$ and there is no such thing as a 'small change of parity'). It is also a *global* symmetry, meaning that the parameter α does not depend on the space-time point x; if it did, so that we had $\alpha \to \alpha(x)$ in (44), the symmetry would be called a *local* one. In the case of (45), the Lagrangian cannot be invariant under such a local phase change because of the $\partial_\mu\phi$ terms, which will produce $\partial_\mu\alpha$ pieces which will not cancel. But, if we include the electromagnetic field, then we can get a Lagrangian which is invariant under local phase transformations (see Section 2.4).

Another piece of jargon we need to introduce is the statement that (44) is a 'U(1)' transformation. The 'U' stands for 'unitary' as in 'unitary matrix'. We can write (44) as $\hat{\phi} \to \mathrm{U}(\alpha)\hat{\phi}$, where the 'matrix' $\mathrm{U}(\alpha)$ has only a single element, i.e. it is a '1×1' matrix. A genuine unitary matrix $\mathbf{U}$ satisfies $\mathbf{U}^\dagger\mathbf{U} = \mathbf{1}$, where $\mathbf{1}$ is the identity matrix and the dagger denotes the Hermitian conjugate. A one-dimensional matrix is of course a single number—in this case a complex number. The 'unitary' condition then reduces to $\mathrm{U}^*\mathrm{U} = 1$, which is to say that U is just a phase factor, as in (44). Such phase factors $e^{i\alpha}$ form a *group*: the product $e^{i\alpha}e^{i\beta}$ of any two of them is also a phase factor, and there is an obvious identity (when $\alpha = 0$) and an inverse (replace α by $-\alpha$). Furthermore, this group is *Abelian*, meaning that it does not matter in which order we multiply any two U's together: $\mathrm{U}(\alpha)\mathrm{U}(\beta) = \mathrm{U}(\beta)\mathrm{U}(\alpha)$. (As we shall see in Section 4, the symmetries of QCD and of the electroweak theory are precisely *non-Abelian generalizations* of (44).) So finally, we say that (44) is a global U(1) transformation, and (45) has a global U(1) symmetry.

The basic theory of such continuous symmetries is supplied by *Noether's theorem*. Because the transformation is continuous, it is good enough to consider an infinitesimal transformation—finite ones can be built up by having lots of little ones. So let us consider an arbitrary $\hat{\mathcal{L}}$ which is invariant under

$$\hat{\phi} \quad \to \quad \hat{\phi}' = \hat{\phi} - i\epsilon\hat{\phi}$$

$$\hat{\phi}^\dagger \quad \rightarrow \quad \hat{\phi}'^\dagger = \hat{\phi}^\dagger + i\epsilon\hat{\phi}^\dagger. \tag{46}$$

The change in $\hat{\mathcal{L}}(\hat{\phi}, \hat{\phi}^\dagger, \partial_\mu\hat{\phi}, \partial_\mu\hat{\phi}^\dagger)$ will then be *zero* (because it is invariant), and this change is

$$\begin{aligned} 0 = \delta\hat{\mathcal{L}} \quad &= \quad \frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi})}\delta(\partial_\mu\hat{\phi}) + \frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi}^\dagger)}\delta(\partial_\mu\hat{\phi}^\dagger) \\ &\quad + \frac{\partial\hat{\mathcal{L}}}{\partial\hat{\phi}}\delta\hat{\phi} + \frac{\partial\hat{\mathcal{L}}}{\partial\hat{\phi}^\dagger}\delta\hat{\phi}^\dagger. \end{aligned} \tag{47}$$

This is a bit like the manipulations leading up to the derivation of the Euler–Lagrange equation in Section 3.1, but now the changes $\delta\hat{\phi}$ and $\delta\hat{\phi}^\dagger$ have nothing to do with space–time trajectories—they mix up the two fields via (46). However, we can *use* the equations of motion for $\hat{\phi}$ and $\hat{\phi}^\dagger$ to rewrite $\delta\hat{\mathcal{L}}$ as

$$\begin{aligned} 0 \quad &= \quad \frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi})}\delta(\partial_\mu\hat{\phi}) + \frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi}^\dagger)}\delta(\partial_\mu\hat{\phi}^\dagger) \\ &\quad + \left[\partial_\mu\left(\frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi})}\right)\right]\delta\hat{\phi} + \left[\partial_\mu\left(\frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi}^\dagger)}\right)\right]\delta\hat{\phi}^\dagger. \end{aligned} \tag{48}$$

Since (see similar steps after (40)) $\delta(\partial_\mu\hat{\phi}_i) = \partial_\mu(\delta\hat{\phi}_i)$, the right-hand side of (48) is just a total divergence, and (48) becomes

$$0 = \partial_\mu\left[\frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi})}\delta\hat{\phi} + \frac{\partial\hat{\mathcal{L}}}{\partial(\partial_\mu\hat{\phi}^\dagger)}\delta\hat{\phi}^\dagger\right]. \tag{49}$$

This means that the quantity inside the [.....] is a 'current' $\hat{j}^\mu$ which is *conserved* in the sense that $\partial_\mu\hat{j}^\mu = 0$.

This is a general result for *any* $\mathcal{L}$ invariant under (46), and it is an example of Noether's theorem (which states that continuous symmetries imply the existence of conserved currents). For our particular case, with the small changes (46), the quantity in the [...] brackets is

$$\begin{aligned} [\ldots] \quad &= \quad (\partial^\mu\hat{\phi}^\dagger).-i\epsilon\hat{\phi} + (\partial^\mu\hat{\phi}).i\epsilon\hat{\phi}^\dagger \\ &= \quad i\epsilon\left((\partial^\mu\hat{\phi})\hat{\phi}^\dagger - (\partial^\mu\hat{\phi}^\dagger)\hat{\phi}\right) \equiv \epsilon\hat{j}^\mu_\phi. \end{aligned} \tag{50}$$

We drop the irrelevant constant parameter ϵ and arrive at the expression for the conserved current following from the symmetry under (46):

$$\hat{j}^\mu_\phi = i\left(\partial^\dagger\hat{\phi})\hat{\phi}^\dagger - (\partial^\mu\hat{\phi}^\dagger)\hat{\phi}\right). \tag{51}$$

What does all this have to do with conserved quantities? Written out in full, the conservation equation $\partial_\mu\hat{j}^\mu_\phi = 0$ is

$$\partial\hat{j}^0_\phi/\partial t + \boldsymbol{\nabla}\cdot\hat{\boldsymbol{j}}_\phi = 0. \tag{52}$$

Integrating this equation over all space, we obtain

$$\frac{d}{dt}\int_{V\to\infty}\hat{j}^0_\phi d^3\boldsymbol{x} + \int_{S\to\infty}\hat{\boldsymbol{j}}_\phi\cdot d\boldsymbol{S} = 0 \tag{53}$$

where we have used the divergence theorem in the second term. Normally the fields die off sufficiently fast at infinity that the surface integral vanishes, and we can therefore deduce that the quantity $\hat{N}_\phi$ is constant in time, where

$$\hat{N}_\phi = \int\hat{j}^0_\phi d^3\boldsymbol{x}; \tag{54}$$

that is, *the volume integral of the $\mu = 0$ component of a symmetry current is independent of time, so its eigenvalues are constants of the motion, i.e. conserved quantum numbers*.

We can calculate $\hat{N}_\phi$ given the field expansion (43). Here we must of course pay attention to the fact that the $\hat{a}$'s and $\hat{b}$'s are mode operators with the commutation relations

$$[\hat{a}(k), \hat{a}^\dagger(k')] = (2\pi)^3\delta(\boldsymbol{k} - \boldsymbol{k'}), \;\; [\hat{b}(k), \hat{b}^\dagger(k')] = (2\pi)^3\delta(\boldsymbol{k} - \boldsymbol{k'}), \tag{55}$$

all other commutators vanishing. Also we are defining the vacuum state $|0\rangle$ as being such that $\hat{a}(k)|0\rangle = \hat{b}(k)|0\rangle = 0$. Now, if we go ahead and calculate $\hat{N}_\phi$ in terms of the $\hat{a}$'s and $\hat{b}$'s from (54), we will get some terms in which the rightmost operator is a creation operator; such terms will not give zero when acting on $|0\rangle$. We want the vacuum to be a state with zero eigenvalue of this conserved quantity, and so we *re-order* the expression for $\hat{N}_\phi$, using the commutation relations, so as to arrive at a form in which all $\hat{a}$'s and $\hat{b}$'s appear to the right of all $\hat{a}^\dagger$'s and $\hat{b}^\dagger$'s (this is called 'normal ordered form'—note that we need to do this with the Hamiltonian also!). We discard (infinite) constant contributions arising from the δ-functions on the RHS of (55). Having done this, we find

$$\hat{N}_\phi = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}[\hat{a}^\dagger(k)\hat{a}(k) - \hat{b}^\dagger(k)\hat{b}(k)], \tag{56}$$

while the Hamiltonian in normally ordered form is

$$\hat{H}_\phi = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}k_0[\hat{a}^\dagger(k)\hat{a}(k) + \hat{b}^\dagger(k)\hat{b}(k)]. \tag{57}$$

So $\hat{N}_\phi$ counts 1 for every 'a' and -1 for every 'b' particle in a state (remember that things like '$\hat{a}^\dagger\hat{a}$' are just number operators), while $\hat{H}_\phi$ counts $+k_0$ for every 'a' and also $+k_0$ for every 'b'. The interpretation then is that free a's and b's of momentum $\boldsymbol{k}$ have the same energy $\sqrt{m^2 + \boldsymbol{k}^2}$, but carry opposite values of the conserved quantum number N_ϕ, which is the eigenvalue of $\hat{N}_\phi$. This is why we interpret $\hat{b}^\dagger$ as the creation operator of an anti-a.

3.3 Fermions

The first step towards getting nearer to the SM is to introduce the quantized Dirac field, which is needed for spin-1/2 particles such as quarks and leptons. The free Dirac equation is

$$i\frac{\partial\psi}{\partial t}(x) = (-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + \beta m)\psi(x) \equiv H_D\psi(x) \tag{58}$$

where the Hamiltonian is thus $H_D = -i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + \beta m$, and $\boldsymbol{\alpha}$ and β are the 4×4 Dirac matrices. As in the scalar case, we shall promote the 'wave function field $\psi(x)$' into a quantum field operator $\hat{\psi}(x)$ with a mode expansion

$$\hat{\psi} = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3\sqrt{2k_0}} \sum_{s=1,2}[\hat{c}_s(k)u(k,s)e^{-ik\cdot x} + \hat{d}_s^\dagger(k)v(k,s)e^{ik\cdot x}], \tag{59}$$

where $k_0 = (m^2 + \boldsymbol{k}^2)^{1/2}$. Note: (i) $\hat{\psi} \neq \hat{\psi}^\dagger$—it is a complex Dirac (spinor) field: as with the complex scalar field, this has to do with the fact that its quanta carry a conserved number which distinguishes particle quanta from antiparticle quanta; (ii) u and v are 4-component spinors of positive and negative 4-momentum respectively, such that

$$(\not{k} - m)u(k,s) = 0, \;\; (\not{k} + m)v(k,s) = 0 \tag{60}$$

where $\not{k} = \gamma^0 k^0 - \boldsymbol{\gamma}\cdot\boldsymbol{k}$ and $\gamma^0 = \beta, \boldsymbol{\gamma} = \beta\boldsymbol{\alpha}$; (iii) there are two independent spinors u (and two independent v's) for given k, corresponding to the two possible spin states for a spin-1/2 particle, labelled by 's'.

We have written (59) in a form which mimics the complex spin-0 case, suggesting that the $\hat{c}$'s are mode annihilation operators and the $\hat{d}^\dagger$'s are mode creation operators. That is, we expect the vacuum to be such that $\hat{c}_s(k)|0\rangle = 0 = \hat{d}_s(k)|0\rangle$, and that particle states will be formed by applying $\hat{c}_s^\dagger$'s and $\hat{d}_s^\dagger$'s to $|0\rangle$. However, while this seems fine for single particle states, we know very well that a state such as

$$|k_1, s_1; k_2, s_2\rangle \propto \hat{c}_{s_1}^\dagger(k_1)\hat{c}_{s_2}^\dagger(k_2)|0\rangle \tag{61}$$

has to be *antisymmetric* under interchange of the labels $(k_1, s_1) \leftrightarrow (k_2, s_2)$: in particular, the state must be zero (fail to exist) if $k_1 = k_2$ and $s_1 = s_2$ (the Pauli exclusion principle). So these mode operators cannot be just like the spin-0 ones.

The solution to this dilemma is simple but radical: for fermions, commutators are replaced by anticommutators! If two different $\hat{c}$'s anticommute, then

$$\hat{c}_{s_1}^\dagger(k_1)\hat{c}_{s_2}^\dagger(k_2) + \hat{c}_{s_2}^\dagger(k_2)\hat{c}_{s_1}^\dagger(k_1) = 0 \tag{62}$$

so that we have the desired antisymmetry

$$|k_1, s_1; k_2, s_2\rangle = -|k_2, s_2; k_1, s_1\rangle. \tag{63}$$

In general we postulate

$$\begin{aligned} \{\hat{c}_{s_1}(k_1), \hat{c}_{s_2}^\dagger(k_2)\} &= (2\pi)^3\delta^3(\boldsymbol{k}_1 - \boldsymbol{k}_2)\delta_{s_1 s_2} \\ \{\hat{c}_{s_1}(k_1), \hat{c}_{s_2}(k_2)\} &= \{\hat{c}_{s_1}^\dagger(k_1), \hat{c}_{s_2}^\dagger(k_2)\} = 0 \end{aligned} \tag{64}$$

and similarly for the $\hat{d}$'s and $\hat{d}^\dagger$'s. The factor in front of the δ function depends on the convention for normalizing Dirac wavefunctions.

Why does it have to be this way? This is a deep question and has a (rather technical) answer in the famous 'spin-statistics theorem' of quantum field theory. One can get some idea of what goes wrong if we use commutators for fermion modes, by considering the Hamiltonian operator which is

$$\hat{H}_D = \int \hat{\psi}^\dagger(x)(-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + \beta m)\hat{\psi}(x)d^3\boldsymbol{x}. \tag{65}$$

If we place the expansion (59) into (65) we find (after quite a lot of algebra)

$$\hat{H}_D = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}k_0 \sum_{s=1,2}[\hat{c}_s^\dagger(k)\hat{c}_s(k) - \hat{d}_s(k)\hat{d}_s^\dagger(k)]. \tag{66}$$

As with $\hat{H}_\phi$ and $\hat{N}_\phi$ for the scalar field, we would want to re-order the last term in (66) so as to ensure $\hat{H}_D|0\rangle = 0$. But if we do this assuming ordinary commutation relations for the $\hat{d}$'s, we get

$$\hat{H}_D = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}k_0 \sum_{s=1,2}[\hat{c}_s^\dagger(k)\hat{c}_s(k) - \hat{d}_s^\dagger(k)\hat{d}_s(k)]. \tag{67}$$

The problem with (67) is that, although indeed $\hat{H}_D|0\rangle = 0$, there are states with *negative* energy!—namely states with any number of d-quanta (because of the minus sign in front of the number operator $\hat{d}^\dagger\hat{d}$). On the other hand, if we re-order the $\hat{d}\hat{d}^\dagger$ term using *anticommutation* relations, we convert the $-$ sign in (67) into a + sign, and all is well.

We can also see the same mechanism at work if we enquire about a conserved *fermion number*. The Dirac Lagrangian is

$$\hat{\mathcal{L}}_D = \hat{\bar{\psi}}(x)(i\gamma\cdot\partial - m)\hat{\psi}(x) \tag{68}$$

where $\hat{\psi}$ and $\hat{\psi}^\dagger$ are independent degrees of freedom (the E–L equation for $\hat{\bar{\psi}}$ is just the Dirac equation $(i\gamma\cdot\partial - m)\hat{\psi} = 0$). The Lagrangian (68) is plainly invariant under the global U(1) transformation

$$\hat{\psi}(x) \to \hat{\psi}'(x) = e^{-i\alpha}\hat{\psi}(x). \tag{69}$$

The corresponding (Noether) symmetry current can be found by following the standard steps in Noether's theorem of §3.2, and is

$$\hat{N}^\mu_\psi = \hat{\bar{\psi}}(x)\gamma^\mu\hat{\psi}(x). \tag{70}$$

The associated symmetry operator is

$$\hat{N}_\psi = \int \hat{N}^0_\psi(x)d^3\boldsymbol{x} = \int \hat{\psi}^\dagger(x)\hat{\psi}(x)d^3\boldsymbol{x}, \tag{71}$$

which is just the usual Dirac number density, integrated over $\boldsymbol{x}$. If we now calculate $\hat{N}_\psi$ from (71), we find

$$\hat{N}_\psi = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\sum_{s=1,2}[\hat{c}^\dagger_s(k)\hat{c}_s(k) + \hat{d}_s(k)\hat{d}^\dagger_s(k)]. \tag{72}$$

The first term is fine, but if we re-order the second to '$\hat{d}^\dagger\hat{d}$' so that $\hat{N}_\psi|\ 0\rangle = 0$, we will be counting +1 for both c's and d's. We clearly need, again, to use *anticommutators*, so that $\hat{N}_\psi \sim \hat{c}^\dagger\hat{c} - \hat{d}^\dagger\hat{d}$, which counts +1 for each c (particles), and -1 for each d (antiparticles).

We also need the Dirac propagator $\langle 0|T(\hat{\psi}(x_1)\hat{\bar{\psi}}(x_2))|0\rangle$. This may be compared with the analogous propagator for the complex scalar field, namely $\langle 0|T(\hat{\phi}(x_1)\hat{\phi}^\dagger(x_2))|0\rangle$—see problem P2.3. But note that in the Dirac case, each of $\hat{\psi}$ and $\hat{\psi}^\dagger$ carries an independent spinor index (telling which of the four components it is), so the Dirac propagator is a 4×4 *matrix* in this spinor space. For the Feynman rule appropriate to a propagating fermion we need the momentum space version, as usual. In the scalar case, the propagator is proportional to $1/(q^2 - m^2)$ where q is the momentum carried by the internal particle and m is its mass. The 'poor man's' way of getting this is to take the equation of motion for a free scalar particle (the KG equation)

$$(\partial_t^2 - \boldsymbol{\nabla}^2 + m^2)\phi(x) = 0 \tag{73}$$

and consider a plane wave solution (4-momentum eigenfunction) of the form

$$\phi = A\exp(-iq^0t + i\boldsymbol{q}\cdot\boldsymbol{x}) = A\exp(-iq\cdot x) \tag{74}$$

giving

$$(-(q^0)^2 + \boldsymbol{q}^2 + m^2)A = (-q^2 + m^2)A = 0 \tag{75}$$

and the propagator is basically the *inverse* of the expression (....) multiplying A in (75), namely $(-q^2 + m^2)^{-1}$. In the Dirac case, an analogous plane wave solution has the form

$$\psi = \exp(-iq\cdot x)u, \tag{76}$$

where u is a 4-component spinor. Inserting (76) into (58) we find

$$(\not{q} - m)u = 0 \tag{77}$$

as in (60), and the inverse of the LHS of (77) is $(\not{q} - m)^{-1}$ (remember that $\not{q}$ is a matrix!). The actual answer is

(iii) a factor $i/(\not{q} - m)$ for an internal fermion line carrying 4-momentum q.

3.4 Local U(1) phase invariance (U(1) gauge theory): QED

Consider the Dirac Lagrangian

$$\hat{\mathcal{L}}_{\rm D} = \bar{\hat{\psi}}(i\gamma\cdot\partial - m)\hat{\psi}. \tag{78}$$

It is certainly invariant under $\hat{\psi} \to e^{-i\alpha}\psi$ with *constant* α, which is a global U(1) symmetry associated with conservation of the number of ψ-fermions, as we have seen. Let us explore the possibility of invariance under the *local* phase transformation

$$\hat{\psi}(x) \to e^{-i\hat{\alpha}(x)}\hat{\psi}(x) \tag{79}$$

where $\hat{\alpha}(x)$ is a scalar quantum field. Clearly $\hat{\mathcal{L}}_{\rm D}$ is not invariant under (79): it changes by

$$\delta\hat{\mathcal{L}}_{\rm D} = \bar{\hat{\psi}}(x)\gamma^\mu\hat{\psi}(x)\partial_\mu\hat{\alpha}(x). \tag{80}$$

Now, in classical electrodynamics, the way in which electromagnetic interactions are introduced in the Hamiltonian formulation of dynamics is via the replacement of the momentum variable p^μ by $p^\mu - eA^\mu$, where $e(>0)$ is the particle's charge and $A^\mu = (V, \boldsymbol{A})$ is the 4-vector of electromagnetic potentials V and $\boldsymbol{A}$ such that $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$ and $\boldsymbol{E} = -\boldsymbol{\nabla}V - \partial\boldsymbol{A}/\partial t$. In quantum mechanics, we follow the same prescription, but now $p^\mu \to \hat{p}^\mu = i\partial^\mu$ and electromagnetism is introduced via $i\partial^\mu \to i\partial^\mu - eA^\mu$, or

$$\partial^\mu \to \partial^\mu + ie\hat{A}^\mu \equiv \hat{D}^\mu. \tag{81}$$

Applying this prescription to $\hat{\mathcal{L}}_{\rm D}$, we generate an *interaction*

$$\hat{\mathcal{L}}_{\rm int} = -e\bar{\hat{\psi}}\gamma^\mu\hat{\psi}\hat{A}_\mu. \tag{82}$$

Now, if $\hat{A}_\mu$ were also to change by exactly the rule

$$\hat{A}_\mu \to \hat{A}_\mu + \frac{1}{e}\partial_\mu\hat{\alpha} \tag{83}$$

when $\hat{\psi}$ changes by (79), the term (80) will be cancelled and the complete Lagrangian $\hat{\mathcal{L}}_{\rm D} + \hat{\mathcal{L}}_{\rm int}$ would be *locally* U(1) invariant.

Of course, this is indeed the case. The electromagnetic potentials *are* arbitrary up to 'gauge transformations' of the form (83) (consider for example just the 3-vector part: $\hat{\boldsymbol{A}} \to \hat{\boldsymbol{A}} + \frac{1}{e}\boldsymbol{\nabla}\hat{\alpha}$, and $\hat{\boldsymbol{B}} = \boldsymbol{\nabla}\times\hat{\boldsymbol{A}}$ remains the same because curl grad =0). So the combined transformations

$$\begin{aligned}\hat{\psi}(x) &\to e^{-i\hat{\alpha}(x)}\hat{\psi}(x)\\ \hat{A}_\mu(x) &\to \hat{A}_\mu + \frac{1}{e}\partial_\mu\hat{\alpha}(x)\end{aligned} \tag{84}$$

are what we mean by a U(1) gauge transformation. *Note* that the interaction is the 4-dimensional dot product of the gauge field $\hat{A}_\mu(x)$ and the 'global U(1) symmetry current' $\bar{\hat{\psi}}\gamma^\mu\hat{\psi}$.

Like our other quantum fields, $\hat{A}^\mu(x)$ has a mode expansion:

$$\hat{A}^\mu(x) = \sum_{\lambda=0}^{3}\int\frac{d^3\boldsymbol{k}}{(2\pi)^3\sqrt{2\omega}}[\epsilon^\mu(k,\lambda)\hat{\alpha}_\lambda(k)e^{-ik\cdot x} + \epsilon^{*\mu}(k,\lambda)\hat{\alpha}^\dagger_\lambda(k)e^{ik\cdot x}] \tag{85}$$

where $\epsilon^\mu(k,\lambda)$ is the 'polarization vector' of the plane wave solution ($\lambda = 0,1,2,3$). $\hat{A}^\mu$ is real (because the photon is its own antiparticle), and $\epsilon^\mu(k,\lambda)$ is a 'spin-1 analogue' of the spinor $u(p,s)$ for the Dirac field.

But this '$\hat{A}^\mu$' is itself a dynamical field, of course. What is *its* Lagrangian? To answer this, we need to find an $\hat{\mathcal{L}}_A$ such that, if that was all we had, the E–L equations of motion would give us the

free-space (source-free) Maxwell equations. Now Maxwell's equations are for the field strengths $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{B}}$, not the potentials, so they are automatically unchanged under the transformation (83)—that is, they are gauge invariant. This suggests that we need to use the gauge invariant object

$$\hat{F}_{\mu\nu} = \partial_\mu \hat{A}_\nu - \partial_\nu \hat{A}_\mu \tag{86}$$

to build our $\hat{\mathcal{L}}_A$ (it is easy to check that $\hat{F}_{\mu\nu}$ is invariant under (83)). Indeed, the Maxwell Lagrangian is

$$\hat{\mathcal{L}}_A = -\frac{1}{4}\hat{F}_{\mu\nu}\,\hat{F}^{\mu\nu}. \tag{87}$$

How do we know? By verifying that indeed the E–L equations for $\hat{A}_\mu$ following from $\hat{\mathcal{L}}_A$ *are* the free-space Maxwell equations (warning: this needs some patience to do correctly, first time!).

So actually we are now in possession of the QED Lagrangian

$$\hat{\mathcal{L}}_{QED} = \bar{\hat{\psi}}(i\,\hat{\not\partial} - m)\hat{\psi} - e\bar{\hat{\psi}}\gamma^\mu\hat{\psi}\hat{A}_\mu - \frac{1}{4}\hat{F}_{\mu\nu}\,\hat{F}^{\mu\nu} \equiv \bar{\hat{\psi}}(i\,\not{D} - m)\hat{\psi} - \frac{1}{4}\hat{F}_{\mu\nu}\,\hat{F}^{\mu\nu} \tag{88}$$

for one fermion of charge e and mass m. *It is invariant under local U(1) transformations—i.e. it is gauge invariant*. What are the Feynman rules? We have the fermion propagator: we need the interaction vertex, and the $\hat{A}^\mu$ (photon) propagator. First, the vertex. Remember that '$\mathcal{L} = T - V$', so the interaction Hamiltonian is

$$\hat{H}' = \int e\bar{\hat{\psi}}\gamma^\mu\hat{\psi}\hat{A}_\mu d^4x. \tag{89}$$

In perturbation theory we always get '$-i\hat{H}'$'. So a lowest order matrix element will be

$$\langle f| - ie\int \bar{\hat{\psi}}\gamma^\mu\hat{\psi}\hat{A}_\mu d^4x|i\rangle. \tag{90}$$

Just as in the 'ABC' case, the amplitude for the elementary building block '$e^- \to e^- + \gamma$' will be just

(iv) $ie\gamma^\mu$

with appropriate factors for an incoming fermion (a u spinor), an outgoing fermion (a $\bar{u}$ spinor), and the γ (ϵ_μ for an ingoing γ, ϵ^*_μ for an outgoing one).

The only other thing we need is the photon propagator, and here we hit an unpleasant snag, which should not be concealed. Let us try to follow the 'poor man's' way of getting propagators in this case. We start with the E-L equation of motion for the A^μ field, which turns out to be

$$\Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = 0 \tag{91}$$

(see problem P2.4). Now try plugging in a free particle plane wave solution $A^\nu \sim \exp(iq\cdot x)\epsilon^\nu$. We get

$$(-q^2\delta^\nu_\mu + q^\nu q_\mu)\epsilon^\mu = 0. \tag{92}$$

The propagator should be basically $(q^2\delta^\nu_\mu + q^\nu q_\mu)^{-1}$. But this inverse does not exist! It is obvious that

$$(-q^2\delta^\nu_\mu + q^\nu q_\mu)q^\mu) = 0 \tag{93}$$

so that treated as a matrix it has a zero eigenvalue; hence its determinant must vanish, and its inverse therefore will not exist.

The propagator *should* be something like $\langle 0|T(\hat{A}^\mu(x_1)\hat{A}^\nu(x_2))|0\rangle$, but as we have seen the $\hat{A}^\mu$'s are not unique, and can be altered by a gauge transformation (83). So *the propagator is in fact gauge dependent*, not a unique quantity, and that is why the naive poor man's approach failed. In classical electromagnetic theory, one 'fixes the gauge', for example by imposing the condition $\partial_\mu A^\mu = 0$, which

reduces (91) to $\Box A^\nu = 0$, and then the plane wave solution gives $-q^2\epsilon^\mu = 0$ and the propagator $\sim 1/q^2$ (as expected!). But in general we must acknowledge the gauge dependence. A standard form for the γ propagator is

(v) a factor $i[-g^{\mu\nu} + (1-\xi)q^\mu q^\nu/q^2]/q^2$ for an internal photon line carrying 4-momentum q, where ξ is a 'gauge parameter' ($\xi = 1$ gives the simple $1/q^2$ form).

Results for physical quantities will always be independent of ξ (i.e. will be gauge invariant), but it is not so simple to give a general proof of this.

Problems for Lecture 2

P2.1 The 'action' in classical mechanics is defined by

$$S = \int_{t_1}^{t_2} [\frac{1}{2}m(\dot{x}(t))^2 - V(x(t))]dt.$$

Consider one-dimensional motion under gravity with $V = -mgx(t)$. Evaluate S for $t_1 = 0, t_2 = T$, for three alternative trajectories: (a) $x(t) = at$; (b) $x(t) = \frac{1}{2}gt^2$ (the Newtonian one); and (c) $x(t) = bt^3$. [Take care to choose a and b so that all trajectories end at the same point.]

P2.2 The classical action is

$$S = \int_{t_1}^{t_2} L[x(t), \dot{x}(t)]dt$$

where L is the Lagrangian. Under an infinitesimal change of trajectory $x(t) \to x(t) + \delta x(t), \dot{x}(t) \to \dot{x}(t) + \frac{d}{dt}\delta x(t)$ the action changes by

$$\delta S = \int_{t_1}^{t_2} [\frac{\partial L}{\partial x}\delta x + \frac{\partial L}{\partial \dot{x}}\frac{d}{dt}\delta x]dt.$$

The classical path is determined from the condition $\delta S = 0$. Show that this implies

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

P2.3 Discuss the interpretation of $\langle 0|T(\hat{\phi}(x_1)\hat{\phi}^\dagger(x_2))|0\rangle$ for both time-orderings.

P2.4 Maxwell's equations are

$$\boldsymbol{\nabla}\cdot\boldsymbol{E} = 0,\ \boldsymbol{\nabla}\times\boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t},\ \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0,\ \boldsymbol{\nabla}\times\boldsymbol{B} = \frac{\partial \boldsymbol{E}}{\partial t}.$$

In quantum mechanics, electromagnetic interactions are introduced via the *potentials* V and $\boldsymbol{A}$ defined by

$$\boldsymbol{E} = -\boldsymbol{\nabla}V = \frac{\partial \boldsymbol{A}}{\partial t},\ \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}.$$

Then $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$ and $\boldsymbol{\nabla}\times\boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}$ are satisfied automatically, while the other two Maxwell equations become

$$(\partial_t^2 - \boldsymbol{\nabla}^2)\boldsymbol{A} + \boldsymbol{\nabla}(\frac{\partial V}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{A}) = 0$$

and

$$(\partial_t^2 - \boldsymbol{\nabla}^2)V - \frac{\partial}{\partial t}(\frac{\partial V}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{A}) = 0.$$

(i) Verify these last two equations.
(ii) Verify that they can be put into a neat *covariant* form by introducing the 4-vector $A^\mu = (V, \boldsymbol{A})$, namely

$$\Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = 0$$

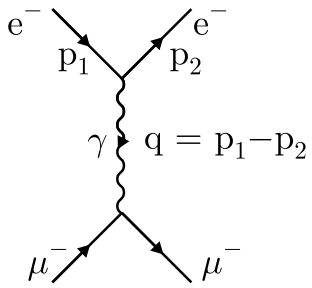


Fig. 4: γ-exchange amplitude in $e^-\mu^- \to e^-\mu^-$.

where $\Box \equiv \partial_t^2 - \boldsymbol{\nabla}^2$, $\partial^\nu \equiv \frac{\partial}{\partial x_\nu}$, $\partial_\mu A^\mu = \frac{\partial V}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{A}$. [Note that $x_i = -x^i$ for $i = 1, 2, 3$; so $\partial^i = \frac{\partial}{\partial x_i} = -\frac{\partial}{\partial x^i} = (-\boldsymbol{\nabla})_{i \text{ component}}$, and $\partial_i = \frac{\partial}{\partial x^i} = (+\boldsymbol{\nabla})_{i \text{ component}}$. So $\partial_\mu A^\mu = \partial_0 A^0 + \partial_i A^i = \frac{\partial V}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{A}$.]

P2.5 Show that

$$[\hat{D}_\mu, \hat{D}_\nu] = ie\hat{F}_{\mu\nu}$$

(see (81)). Hint: in working with such commutators of differential operators, it is best to put in an arbitrary function for the operators to act on, on both sides.

P2.6 A photon mass term in the Lagrangian would give a term proportional to $\hat{A}^\mu \hat{A}_\mu$. Show that this is *not* gauge invariant.

4 ONE-LOOP GRAPHS IN QED: RENORMALIZATION, AND RUNNING COUPLING CONSTANT

See Chapters 10 and 11 of Ref. [1].

Feynman diagrams represent terms in a perturbation theory expansion of physical amplitudes, where the expansion parameter is the relevant 'charge' of the theory—'e' for QED, or more precisely the fine structure constant $\alpha = e^2/4\pi$. The lowest order graphs for any process are always the ones with the fewest vertices, and this means, in fact, that for given external 'legs', each vertex must be joined to only one other vertex by a single internal line (propagator); for example, the γ-exchange amplitude in $e^-\mu^- \to e^-\mu^-$ shown in Fig. 4. Such graphs are called 'tree' graphs.

But tree graphs will only give us the lowest order contribution to the amplitudes. As soon as we go to the next order in perturbation theory, we meet *loops*—for example, those shown in Figs. 5 (a), (b) and (c), which are O(α^2) (four powers of e) diagrams in $e^-\mu^- \to e^-\mu^-$. Admittedly, since $\alpha \sim 1/137$ is quite small, such corrections would seem to be relatively insignificant, perhaps. But, as you all know very well, there are certain quantities (such as the anomalous magnetic moments of the e and the μ) which are known with truly remarkable precision (typically 0.1%), well beyond that represented by the simplest lowest order calculation. More to the point for this school, LEP and SLAC experiments had an accuracy sensitive to one-loop corrections; hence an understanding of this physics is now essential for phenomenology.

As soon as one tries to calculate a loop, in nearly all quantum field theories, one finds that it is infinite! This is pretty disastrous, particularly as loops are supposed to be a small correction to the tree graphs (if the expansion parameter is small, as α is). Thus at once we are faced with the whole business of *renormalization*, which is a systematic procedure for 'taming' these infinities. All three gauge theories of the Standard Model are 'renormalizable', meaning that higher order corrections can in fact be reliably calculated. The remarkable agreement between theory and experiment is impressive confirmation that the rather elaborate theoretical structure of these theories is actually a good model of nature at this scale.

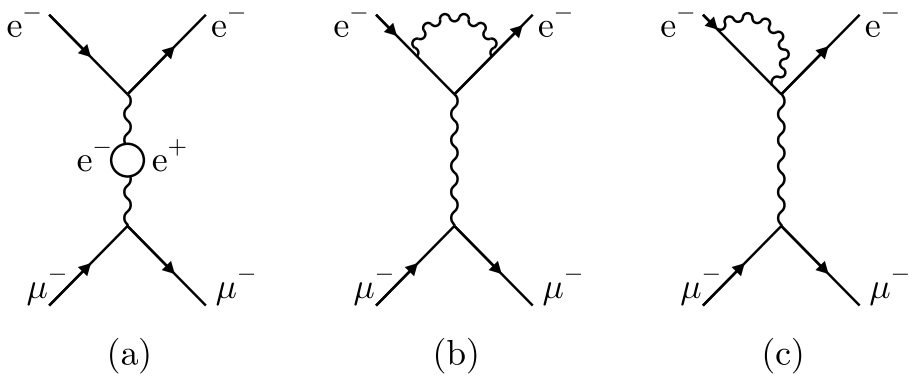


Fig. 5: One-loop corrections to Fig. 4.

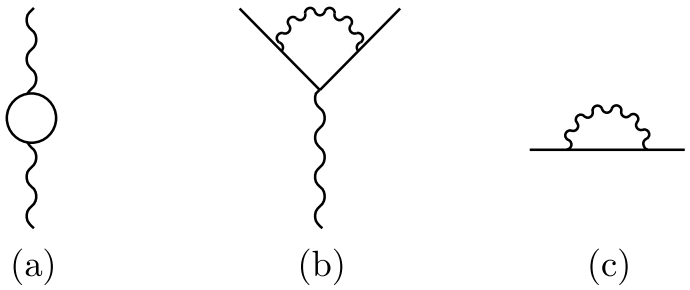


Fig. 6: The loop parts of Fig. 5.

However, the renormalization of non-Abelian gauge theories is too technical for this course, and here I shall sketch how it works for QED only.

The loop bits in Fig. 5 are, in fact, the only divergent one-loop graphs in QED; we redraw them separately in Figs. 6 (a)–(c). Figure 6(a) is clearly a correction to the photon propagator, and is called generically a 'vacuum polarization' graph (see Section 5.3), (b) is a 'vertex correction' and (c) is a correction to a fermion propagator. We are going to concentrate on (a).

4.1 Vacuum polarization and the photon self-energy

We shall use the gauge $\xi = 1$ in which the unmodified photon propagator is $-ig^{\mu\nu}/q^2$. The amplitude for Fig. 6(a) is (omitting Dirac spinor factors for the fermion lines)

$$\frac{-ig^{\nu\rho}}{q^2}\left(i\Pi^{[2]}_{\rho\sigma}(q^2)\right)\frac{-ig^{\sigma\mu}}{q^2} \tag{94}$$

where

$$i\Pi^{[2]}_{\rho\sigma}(q^2) = (-1)(-ie)^2\mathrm{Tr}\int\frac{d^4k}{(2\pi)^4}\frac{i}{\not q+\not k-m}\gamma_\rho\frac{i}{\not k-m}\gamma_\sigma. \tag{95}$$

Note: (i) When we attach external legs to Fig. 6(a), as in Fig. 5(a), 'q' will be determined in terms of the 4-momenta of the external particles, but this q is *shared* by the e^+ and e^- in the loop in all possible ways: the e^+ has 4-momentum k, say, in the direction indicated, and the e^- has $q + k$, but nothing determines k—it has to be *integrated over*. (ii) The (-1) factor has to be included for all closed fermion loops, as does the Tr (which means 'take the trace—i.e. sum the diagonal elements—of the Dirac matrix product').

The $\int d^4x$ in (95) extends over the (presumably) infinite 4-D 'volume'; in particular, all components of k can go to infinity. So a crude 'counting of powers' seems to show that (95) will diverge

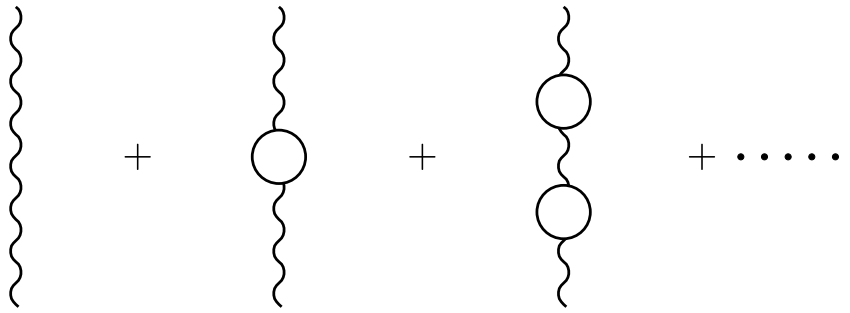


Fig. 7: Sum of vacuum polarization 'bubble' insertions.

as

$$\int d^4k/k^2 \sim \int k^3 dk/k^2 \sim \int k dk \sim \Lambda^2 \tag{96}$$

if we 'cut off' the integral at an upper limit Λ. This would be a (divergent) *constant* contribution, multiplying $g_{\rho\sigma}$ to get the indices right. What would such a constant loop correction mean, in this case? Suppose we consider a whole series of such 'insertions', as shown in Fig. 7—which is, in fact, a geometric series of the form '$\frac{1}{a} + \frac{1}{a}b\frac{1}{a} + \frac{1}{a}b\frac{1}{a}b\frac{1}{a}\ldots$', summing to $\frac{1}{a(1-b/a)} = \frac{1}{a-b}$. In the present case, then, this would mean that a constant part of $\Pi^{[2]}_{\rho\sigma}$ will correct the propagator (after summing) to something of the form $(q^2 - \text{const})^{-1}$—in other words, the photon will apparently acquire a mass!

Actually, such insertions into propagators usually do have the effect of shifting the mass of the particle in question, and they are generically called 'self-energies' (e.g. Fig. 6(c) is a fermion self-energy, which will indeed modify the original fermion mass). But the real photon is massless! We know this to a very high accuracy experimentally. Theoretically, this is fundamentally related to *gauge invariance*—see problem P2.6! So, provided we introduce the cut-off in a *gauge-invariant* way, it turns out that this apparent Λ^2 divergence of (95) is not there after all. Instead, what one finds is that

$$i\Pi^{[2]}_{\rho\sigma}(q^2) = i(q^2 g_{\rho\sigma} - q_\rho q_\sigma)\Pi^{[2]}_\gamma(q^2) \tag{97}$$

where $\Pi^{[2]}_\gamma(q^2)$ is a Lorentz scalar, and is given by an integral which diverges more 'weakly', namely as $\ln\Lambda$. Note that the dimensions of $\Pi^{[2]}_{\rho\sigma}(q^2)$ are M^2: in the 'naive' cut-off approach this was visible in the Λ^2, whereas in (97) quadratic factors of q appear, and this is why the divergence can only be logarithmic. These factors ensure that

$$q^\rho \Pi^{[2]}_{\rho\sigma} = q^\sigma \Pi^{[2]}_{\rho\sigma} = 0 \tag{98}$$

(assuming $\Pi^{[2]}_\gamma$ is finite!); this guarantees that the ξ-dependent part of the propagator (rule(v)) disappears—i.e. the result is gauge invariant, as required.

When all the bubbles are added up, and bits proportional to q are omitted because of (98) (gauge invariance), one finds the net result that the photon propagator is modified according to

$$\frac{-ig_{\mu\nu}}{q^2} \to \frac{-ig_{\mu\nu}}{q^2\left(1 - \Pi^{[2]}_\gamma(q^2)\right)}. \tag{99}$$

What is the physics of this? When (99) appears inside a scattering graph such as Fig. 5(a), we would still be able to say that the (corrected) exchanged photon had zero mass, since near the 'mass shell' point $q^2 = 0$ (99) does indeed behave like (a constant times) the massless propagator $1/q^2$, provided that $\Pi^{[2]}_\gamma(q^2 = 0)$ is finite.

Discussion Point: What happens if $\Pi^{[2]}_\gamma(q^2 = 0)$ itself has a term like A/q^2? and how might this happen?

On the other hand, the propagator will have a peculiar *normalization*: it will be

$$\sim \left(\frac{1}{1 - \Pi_\gamma^{[2]}(0)} \right) \cdot \frac{-ig_{\mu\nu}}{q^2} \tag{100}$$

for $q^2 \approx 0$ instead of the familiar $-ig_{\mu\nu}/q^2$. Why is this? The propagator in the *free* case was the Fourier transform of $\langle 0|T(A_\mu(x_1)A_\nu(x_2))|0\rangle$. Take one time-ordering, say $\langle 0|A_\mu(x_1)A_\nu(x_2)|0\rangle$, and insert a complete set of free states via '$\sum_n |n\rangle\langle n| = 1$':

$$\sum_n \langle 0|A_\mu(x_1)|n\rangle\langle n|A_\nu(x_2)|0\rangle. \tag{101}$$

The *only* state 'n' which can contribute is the state of one free photon—and indeed we know that matrix elements of the form $\langle 0|\text{field operator}|\text{particle state}\rangle$ are always just the corresponding *wavefunction*. But now consider the interacting case. Here the full propagator is $\langle\Omega|T(A_\mu(x_1)A_\nu(x_2))|\Omega\rangle$ where $|\Omega\rangle$ is the exact 'interacting' vacuum. Insert a complete set of interacting states $\sum_n |\bar{n}\rangle\langle\bar{n}| = 1$: then the analogue of (101) is

$$\sum_n \langle\Omega|A_\mu(x_1)|\bar{n}\rangle\langle\bar{n}|A_\nu(x_2)|\Omega\rangle \tag{102}$$

and now the crucial point is that in addition to the one-photon state in $|\bar{n}\rangle$ there will also be a whole lot of *other* states to which the photon can couple—for instance, precisely the e^+e^- state in our vacuum polarisation graph! This must mean that the $|1\gamma\rangle$ state cannot any longer, by itself, produce all of the '1' in the completeness sum. So the 'strength' of the matrix element $\langle 1\gamma|A_\mu(x)|\Omega\rangle$ cannot be unity (in the normalization we are adopting, like problem P1.4).

To take account of this 'diminished single particle strength', we write

$$\langle\gamma, k, \lambda|A_\mu(x)|\Omega\rangle = \sqrt{Z_3}\epsilon_\mu^*(\lambda)e^{ik.x} \tag{103}$$

where Z_3 is called the wavefunction renormalization constant. This will mean that the interacting propagator has the form

$$\begin{aligned} &\text{F.T. of}\langle\Omega|T(A_\mu(x_1)A_\nu(x_2))|\Omega\rangle \\ &= \tfrac{-iZ_3 g_{\mu\nu}}{q^2} + \text{contributions from non single particle states,} \end{aligned} \tag{104}$$

for $q^2 \approx 0$. So we can identify

$$Z_3 = \frac{1}{1 - \Pi_\gamma^{[2]}(0)}. \tag{105}$$

This is the interpretation of the change in normalization of the photon propagator.

This is all innocent-sounding enough ... but of course $\Pi_\gamma^{[2]}(0)$ depends on Λ and is *divergent* as the cut-off $\Lambda \to \infty$. To bury this divergence, which after all is occurring as a multiplicative factor in the wavefunction, we introduce the 'physical' (*renormalized*) photon field operator $A_{\mu,\text{ph}}$ defined by

$$A_{\mu,\text{ph}}(x) = \frac{1}{\sqrt{Z_3}} A_\mu(x) \tag{106}$$

for which the propagator will be of the expected form

$$\text{F.T. of}\langle\Omega|T(A_{\mu,\text{ph}}(x_1)A_{\nu,\text{ph}}(x_2))|\Omega\rangle \approx \frac{-ig_{\mu\nu}}{q^2} + \text{multiparticle bits} \tag{107}$$

for $q^2 \approx 0$. Formally this will work even if $\Lambda \to \infty$; the *physical* matrix elements are OK. Note that $Z_3 = Z_3(\Lambda)$, from (105), since $\Pi_\gamma^{[2]}$ depends on Λ.

Discussion Point: Do we actually envisage $\Lambda \to \infty$, really?

Now let us tidy up. Our results so far tell us that the *renormalized* γ-propagator is $Z_3^{-1}\times$ the one we have been calculating to $O(\alpha)$, that is

$$\frac{1}{Z_3}\cdot\frac{-ig_{\mu\nu}}{q^2\left(1-\Pi_\gamma^{[2]}(q^2,\Lambda)\right)} \tag{108}$$

where we now indicate the Λ dependence explicitly. Now

$$Z_3(\Lambda)=[1-\Pi_\gamma^{[2]}(0,\Lambda)]^{-1}\approx[1+\Pi_\gamma^{[2]}(0,\Lambda)] \tag{109}$$

since $\Pi^{[2]}\sim\alpha$ and we are doing a systematic order-by-order perturbative approach. So (108) becomes

$$\approx\frac{-ig_{\mu\nu}}{q^2\left(1-\Pi_\gamma^{[2]}(q^2,\Lambda)+\Pi_\gamma^{[2]}(0,\Lambda)\right)} \tag{110}$$

again dropping the $O(\alpha^2)$ term $\Pi_\gamma^{[2]}(q^2)\Pi_\gamma^{[2]}(0)$. So finally our renormalized propagator is

$$\frac{-ig_{\mu\nu}}{q^2\left(1-\bar\Pi_\gamma^{[2]}(q^2)\right)} \tag{111}$$

where

$$\bar\Pi_\gamma^{[2]}(q^2)=\lim_{\Lambda\to\infty}[\Pi_\gamma^{[2]}(q^2,\Lambda)-\Pi_\gamma^{[2]}(0,\Lambda)] \tag{112}$$

is called the 'once-subtracted self-energy', and is *finite* and independent of Λ as $\Lambda\to\infty$. We will come back to (111) in Section 4.3.

4.2 The fermion self-energy and the vertex correction

Let us now briefly examine the other two one-loop divergent graphs, Figs. 6(b) and 6(c), beginning with the latter, the fermion self-energy. In analogy with $\Pi_\gamma^{[2]}$, we call the amplitude for Fig. 6(c) $-i\Sigma^{[2]}(p)$ where

$$-i\Sigma^{[2]}(p)=(-ie)^2\int\gamma^\nu\frac{-ig_{\mu\nu}}{k^2}\frac{i}{\not p-\not k-m}\gamma^\mu\frac{d^4k}{(2\pi)^4}. \tag{113}$$

As in the γ case, when the string of self-energy insertions is summed up, the result is a modified fermion propagator equal to

$$\frac{i}{\not p-m-\Sigma^{[2]}(p)}. \tag{114}$$

As expected, $\Sigma^{[2]}$ as given by (113) diverges: there are four powers of k in the numerator and three in the denominator, so we might expect a divergent term proportional to Λ (note that $\Sigma^{[2]}$ has dimensions of mass, as is also evident from (114)). Actually the leading p-independent divergence is, instead, proportional to $m\ln(\Lambda/m)$. The reason for this is important, and it has interesting generalizations. Suppose that m in the Dirac Lagrangian $\bar\psi(i\not\partial-m)\psi$ were were set equal to zero. Then (see problem P3.1) the two 'left' and 'right' helicity components $\psi_L=\left(\frac{1-\gamma_5}{2}\right)\psi$ and $\psi_R=\left(\frac{1+\gamma_5}{2}\right)\psi$ of the electron field will not be coupled by the QED interaction. It follows that no terms of the form $\bar\psi_L\psi_R$ or $\bar\psi_R\psi_L$ can be generated—and these are just of the 'Dirac mass' type (problem P4.2). Hence no perturbatively-induced fermion mass term can be generated by higher-order electromagnetic interactions, and the $\Sigma^{[2]}$ correction must vanish as $m\to0$. So it must behave as $\sim m\ln(\Lambda/m)$ on dimensional grounds, which gives a logarithmically divergent correction to m in (114), call it $\delta m^{[2]}(\Lambda)$.

We can agree to call the resulting 'on shell point $\not{p} = m + \delta m^{[2]}(\Lambda)$' the *physical* mass $m_{\rm ph}$, such that

$$m_{\rm ph} = m(\Lambda) + \delta m^{[2]}(\Lambda) \tag{115}$$

is independent of Λ as $\Lambda \to \infty$—which of course means that the *original* parameter m has in fact to be Λ-dependent, and in just such a way as to compensate for that of $\delta m^{[2]}$.

There is also a p-dependent logarithmic divergence of the form $\not{p} \ln \Lambda/m$. This can be soaked up in a fermion wavefunction renormalization constant Z_2, analogous to Z_3, and having the same interpretation:

$$\psi_{\rm ph} = \frac{1}{(Z_2)^{\frac{1}{2}}}\psi. \tag{116}$$

In this way the physical fermion propagator is indeed

$$i/(\not{p} - m_{\rm ph}). \tag{117}$$

Finally there is the vertex part shown in Fig. 6(b). In this case, power counting indicates a new logarithmic divergence. We have one more card to play, in order to sweep it up. Consider the QED interaction term

$$-e\bar{\psi}(x)\, \not{A}(x)\psi(x) = -e\bar{\psi}_{\rm ph}\, \not{A}_{\rm ph}\psi_{\rm ph} \cdot Z_2 Z_3^{\frac{1}{2}}. \tag{118}$$

This generates a 'lowest order' vertex (in terms of the physical renormalised fields) equal to $-ie\gamma^\mu Z_2 Z_3^{\frac{1}{2}}$ to which Fig. 6(b) must be added. Now the physical charge $e_{\rm ph}$ is going to be determined experimentally from the Coulomb scattering contribution as $q^2 \to 0$ (the classical limit). Figure 6(b) contributes a logarithmically divergent correction to the charge in this limit, call it $\delta e(\Lambda)$. So, once again, we are going to assume that the 'original' e had a Λ-dependence just right to cancel out the Λ-dependence of the total contribution, leaving a finite Λ-independent physical charge as $\Lambda \to \infty$. We express this formally by introducing the vertex renormalization constant Z_1 such that the physical charge is defined by

$$e_{\rm ph} = Z_2 Z_3^{\frac{1}{2}}(e/Z_1). \tag{119}$$

The interaction (118) then becomes

$$-Z_1 e_{\rm ph}\bar{\psi}_{\rm ph}\, \not{A}_{\rm ph}\psi_{\rm ph}. \tag{120}$$

Now some alarm bells should be ringing! The free Dirac part of the QED Lagrangian is now

$$\bar{\psi}(i\ \not{\partial} - m)\psi = Z_2\bar{\psi}_{\rm ph}(i\ \not{\partial} - m)\psi_{\rm ph} \tag{121}$$

to which we must add (120) (as well as the Maxwell term). But then the result is *not* gauge invariant!—since $\not{\partial}$ does not appear in the gauge invariant combination '$\not{\partial} + ie\ \not{A}$' (see Section 3). For this to work we need a kind of small miracle—the equality

$$Z_1 = Z_2 \tag{122}$$

between two quite different wavefunction renormalization constants. Of course, (122) *is* true; it is a *Ward identity*, and can be proved to follow from the gauge invariance of the original QED Lagrangian.

Relation (122) has a remarkable consequence: the 'rescaling' relation (119) now becomes

$$e_{\rm ph} = \sqrt{Z_3}e \tag{123}$$

showing that the corrections to 'e' associated with the fermion propagator and the vertex cancel out, leaving only the γ-propagator correction. Now this correction is the same whatever the external particles are, in a Feynman graph. So (123) is a statement of 'universality' of radiative corrections: they do not

spoil the gauge invariance of the original Lagrangian, and the ration of e to $e_{\rm ph}$ is independent of the types of external particles. If a set of unrenormalized charges are all equal (or 'universal'), the renormalized ones will be too. Universality survives renormalization—and this is a very big clue as to why the weak interactions have to be described by a gauge theory too, since quarks and leptons do seem to couple in some 'universal' way to W's and Z's: the strong interactions, experienced only by the quarks, do not seem to spoil that, just as—in the electromagnetic case—the charge on a proton is the same as that on a positron.

4.3 The physics of $\bar{\Pi}_\gamma^{[2]}(q^2)$

We shall only be able to offer brief notes:

(i) How does the renormalized γ-propagator affect physical processes? Let us imagine using it in $e^-\mu^- \to e^-\mu^-$ scattering via Fig. 4 with the corrected propagator (111), for instance. Then, the amplitude will be (omitting the spinor factors)

$$(-ie)^2 \frac{-ig_{\mu\nu}}{q^2\left(1-\bar{\Pi}_\gamma^{[2]}(q^2)\right)} \tag{124}$$

where now we have changed the notation so that 'e' means the *physical* charge (which we previously called $e_{\rm ph}$), and m is the physical mass (previously $m_{\rm ph}$). In the static limit $q_0 = 0$, the photon propagator $\sim 1/\boldsymbol{q}^2$ has a simple interpretation—it is the Fourier transform of the $1/r$ Coulomb potential (see 'Point 4' at the end of Section 2). So the form (124) must, in the static limit, represent *corrections* to Coulomb's law. Indeed, with $q_0 = 0$ and evaluating $\bar{\Pi}_\gamma^{[2]}$ for $\boldsymbol{q}^2 \ll m^2$, one finds that (124) becomes, approximately,

$$(-ie^2)\frac{ig_{\mu\nu}}{-\boldsymbol{q}^2}\left(1+\frac{\alpha}{15\pi}\,\boldsymbol{q}^2/m^2\right) \tag{125}$$

$$\sim \frac{e^2}{\boldsymbol{q}^2} + \text{constant}. \tag{126}$$

The $e^2/\boldsymbol{q}^2$ in (126) gives us back the Coulomb $1/r$ in x-space: the Fourier transform of the 'constant' is a δ function. This very short distance correction, affecting only s-states in atomic physics, is responsible for a small (but entirely detectable) contribution to the famous *Lamb shift* between hydrogenic $2^2S_{\frac{1}{2}}$ and $2^2P_{\frac{1}{2}}$ levels. See problem P3.3.

(ii) Without making the low-$\boldsymbol{q}^2$ approximation, the form $\sim \frac{e^2}{\boldsymbol{q}^2}\left(1+\bar{\Pi}_\gamma^{[2]}(q^2)\right)$ indicates that the charged leptons have effectively developed a 'form factor' (or spatial extension, when Fourier transformed) due to radiative corrections. Sharing it equally between the two e's in 'e^2', we can say that the radiatively induced charge form factor is $\mathcal{F}_1(q^2) \approx 1 + \frac{1}{2}\bar{\Pi}_\gamma^{[2]}(q^2)$. Examination of the Fourier transform of this shows that the spatial extension is of order $\sim m^{-1}$, the fermion Compton wavelength.

(iii) An alternative interpretation is in terms of a 'q^2-dependent charge', or 'q^2-dependent α', given by

$$\alpha(q^2) = \alpha\left[1+\bar{\Pi}_\gamma^{[2]}(q^2)\right]. \tag{127}$$

The idea that a charge is q^2-dependent may seem strange at first, but it is analogous to the way in which a charge placed in a polarizable medium can give rise to a *space*-dependent effective charge, due to screening (see Fig. 8). The screening length here is just m^{-1}, the distance over which the e^+e^- pairs can be 'fluctuated' out of the vacuum, and which measures the extension of the radiatively induced form factor. This is why the photon self-energy e^+e^- bubble is called a vacuum polarization graph!

For $|q^2| \gg m^2$, (127) becomes

$$\alpha(q^2) \approx \alpha\left[1+\frac{\alpha}{3\pi}\ln\left(\frac{-q^2}{m^2}\right)\right] \tag{128}$$

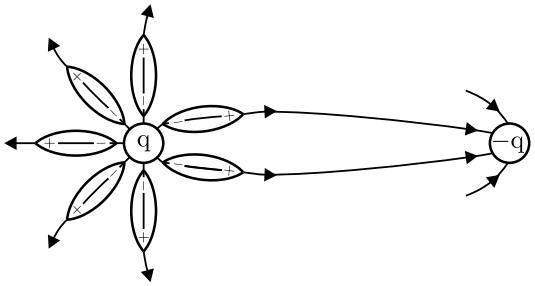


Fig. 8: Screening of a charge in a dipolar medium.

showing that $\alpha(q^2)$ increases at large $-q^2$ (which is short distances, when Fourier transformed), just as indicated in Fig. 8.
(iv) However, a better approximation at large $-q^2$ is to return to the form (124) and write

$$\alpha(Q^2) = \frac{\alpha}{[1 - (\alpha/3\pi)\ln(Q^2/m^2)]} \quad \text{for} \quad Q^2 \gg m^2 \tag{129}$$

where $Q^2 = -q^2$. Equation (129) is the standard 'leading log' expression for the *running coupling constant* in QED. This shows a slow logarithmic increase as Q^2 increases. For example, $\alpha(M_{\rm Z}^2) \sim 1/128.8$, as compared with $\alpha(=\alpha(0)) \sim 1/137$. In QCD, the effect of gluon self-interactions is to make α_s (the QCD analogue of α) *decrease* as Q^2 increases ('asymptotic freedom'). There, the analogous formula is

$$\alpha_{\rm s}(Q^2) = \frac{\alpha_{\rm s}}{\left[1 + \frac{\alpha_{\rm s}}{12\pi}(33 - 2f)\ln(Q^2\mu^2)\right]} \tag{130}$$

where f is the number of fermion–antifermion pairs (in the loops) considered, and μ is a 'renormalization scale'. If $f < 16$, α_s will decrease as Q^2 increases, leaving the quarks weakly interacting at very short distances.

4.4 Renormalizability

We have tried to give some idea of how we can make sense of a theory with divergences. At the one-loop level, some of the steps seemed quite trivial. More generally, however, we can ask: how do we know that we can *go on* soaking up these divergences into redefinitions of 'physical' quantities, as we proceed on to higher order loops? The answer is really rather remarkable: there *are* classes of theory ('renormalizable theories') which are such that *all* divergences, encountered at each successive order in systematic perturbation theory, can be tamed by this procedure of redefining finite physical quantities (and doing wavefunction rescalings), and then re-expressing all amplitudes in terms of these physical quantities. Furthermore, there is a surprisingly simple criterion for telling (almost) which theory is renormalizable and which is not. This criterion has to do with the *dimensionality of the coupling constant* (in units $\hbar = c = 1$)—see problem P3.4.

The result is simply stated: if the dimensionality of the coupling is M^a where $a > 0$, then the theory is 'super-renormalizable' (like the ABC theory—there are fewer divergences than we could in fact deal with, for instance Z_C and the vertex correction are finite); if $a = 0$ (dimensionless) then the theory *may* be renormalizable, and often is (e.g. QED, where the coupling is α); and if $a < 0$, the theory is not renormalizable.

Consider a hypothetical theory, similar to the original four-fermion theory of β-decay, describing interactions between the ν_e and a neutron (assumed pointlike for this purpose). The interaction density is

$$G_F \bar{\psi}_n(x)\psi_n(x)\bar{\psi}_{\nu_e}(x)\psi_{\nu_e}(x). \tag{131}$$

To find the dimensionality of G_F, we need to remember that the mass term in the Dirac Hamiltonian is $m\bar{\psi}\psi$, so that the dimension of a ψ field is $M^{\frac{3}{2}}$. This implies that the dimension of G_F is M^{-2} so that this theory is non-renormalizable. Is this in fact so bad? Consider what happens when we calculate $n+\nu_e \to n+\nu_e$ in perturbation theory. The lowest order ('tree') graph is Fig. 9(a); next is Fig. 9(b); and then at third order Fig. 9(c). Let us count powers in the loop of Fig. 9(b). Since each fermion propagator $\sim k^{-1}$, we expect the graph to diverge as Λ^2. Fine ... what about Fig. 9(c)? Here we have two loops, with

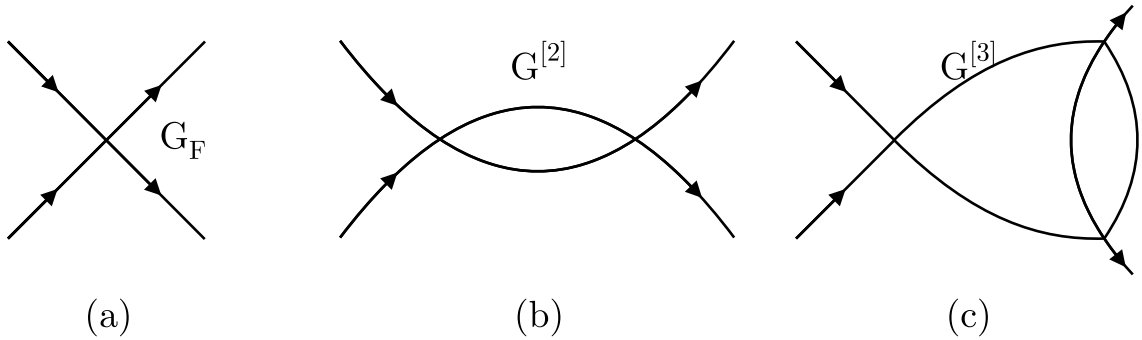


Fig. 9: Contributions to $n + \nu_e \to n + \nu_e$ in perturbation theory, using (131).

therefore eight momentum integrals, and four fermion propagators each contributing only one power of k in the denominator, so it diverges as Λ^4! The first point to note, then, is clearly that as we go up in order of perturbation theory, the divergence gets *worse*. To control the Λ^4 divergence, we would have to 'subtract' the amplitude for Fig. 9(c) *three* times. Each subtraction means that we have to take one parameter from experiment (the amplitude at a certain point, its derivative at that point, its second derivative, etc). Very soon we need more parameters than are appearing in the original Lagrangian (masses, couplings). So simply defining a 'physical' set of Lagrangian parameters will not get us off the hook in this case. A renormalizable theory is one whose infinities can all be tamed by redefinitions of the parameters in the original Lagrangian (plus wavefunction rescalings); if infinities arise which need new parameters (not in the original Lagrangian) to be taken from experiment, then the theory is non-renormalizable.

The reason for this worsening divergence in higher orders in G_F is, of course, related to the dimensionality of G_F. All the amplitudes of Fig. 9 have to have the same dimension, obviously. But since each G_F brings in two powers of a mass 'M' in the denominator, these must be compensated by two powers of momentum in the numerator, making the divergence successively worse.

Is the situation really hopeless? Actually no. We know quite well that people lived with the Fermi theory reasonably happily for years, *until* the advent of high-energy experiments probing weak interactions. The reason can again be found in dimensional analysis. Consider the amplitude for Fig. 9(b), call it $G^{[2]}(s)$, where $s = (p_1 + p_2)^2$. This needs two subtractions to tame it into a finite quantity $\bar{G}_F(s) = G^{[2]} - G^{[2]}(s_0) - (s - s_0)\frac{dG^{[2]}}{ds}\Big|_{s=s_0}$, where s_0 is the point we choose to define our amplitudes at. This means that, expanding $\bar{G}^{[2]}(s)$ about $s = s_0$, we can *calculate* terms of order $(s - s_0)^2$ and higher (the two lowest terms in the expansion have to be taken from experiment). But the worse divergence of Fig. 9(c) (amplitude $G^{[3]}$) would require us to do *three* subtractions before arriving at a finite part we could calculate: in this case, the first calculable bit would be $\sim G_F^3(s - s_0)^3\frac{d^3G^{[3]}}{ds^3}\Big|_{s=s_0}$ —and the process has to be repeated each time we go up an order. Assuming that all the derivatives are about the same order of magnitude, we see that we can get away with using only low order corrections provided $G_F s \ll 1$, i.e.

$$\sqrt{s} \ll \frac{1}{\sqrt{G_F}} \quad . \tag{132}$$

This is an important idea—and in the case of the real Fermi constant ($G_F \sim 1.17\times10^{-5}\mathrm{GeV}^{-2}$), $\frac{1}{\sqrt{G_F}} \sim$

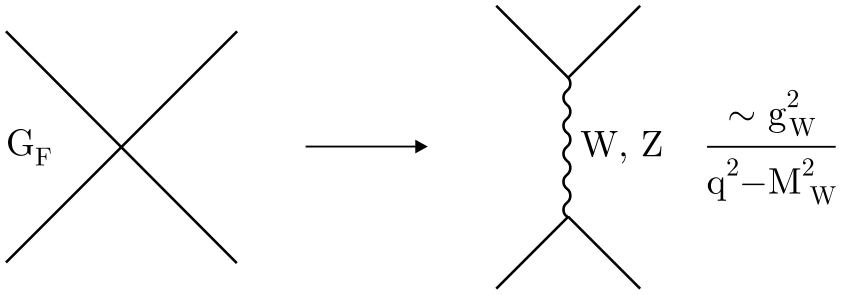


Fig. 10: Relation between four-fermi coupling and Yukawa-like coupling.

300 GeV. So a non-renormalizable theory can be useful at energies well below its 'natural energy scale', as set by the inverse coupling constant; but the nearer we approach this scale, the less predictive the theory will become. And we are, after all, always striving to *reduce* the number of parameters in our theories that have to be taken from experiment.

From this perspective, it may be less of a mystery why renormalizable theories are generally the *relevant* ones at present energies. We may imagine that a 'true' theory exists at some enormously high scale Λ (the Planck scale?) which, though not itself a local quantum field theory, can be written out in terms of all possible fields and their couplings, as allowed by the operative symmetry principles. Our particular renormalizable *subset* of these theories then emerges as a low-energy effective theory, due to the strong suppression of the non-renormalizable terms (which are damped like (s/Λ^2) to some power).

Nonrenormalizable theories may be physically detectable at low energies if they involve processes that would be otherwise forbidden. For example, the fact that (as far as we know) neutrinos have neither electromagnetic nor strong interactions, but only weak ones, allowed the four-fermi interaction to be detected—but amplitudes were suppressed by powers of s/M_W^2 relative to electromagnetic ones, and this is precisely why it was 'weak'! As we shall discuss later, the four-fermi model is superseded in the Standard Model by a Yukawa-type theory involving exchanges of $W^{\pm}, Z^0$ (see Fig. 10). For $q^2 \ll M_W^2$, $G_F \sim g_W^2/M_W^2$, explaining the origin of the M^{-2} dimensionality of G_F, *and* telling us the actual scale, in this case. Thus this theory changes from being an effective non-renormalizable four-fermion theory at very low energies, to being an effective renormalizable one at $q^2 \sim M_W^2$.

Problems for Lecture 3

P3.1 For a Dirac field $\psi(x)$, define $\psi_R = \left(\frac{1+\gamma_5}{2}\right)\psi, \psi_L = \left(\frac{1-\gamma_5}{2}\right)\psi$. Show that

$$\bar{\psi}_L \gamma^\mu \psi_R = 0 \ ,$$

where $\bar{\psi}_L = \psi_L^\dagger \gamma_0$.

P3.2 Rewrite $m\bar{\psi}\psi$ in terms of the ψ_R and ψ_L fields, and deduce that electromagnetic interactions cannot generate such a 'Dirac' mass in perturbation theory.

P3.3 Coulomb's law is corrected by the vacuum polarization (e^+e^-) to

$$-\left\{\frac{\alpha}{r} + \frac{4\alpha^2}{15m^2}\delta^3(\boldsymbol{r})\right\}$$

where m is the electron mass. Treating the δ function piece as a perturbation on the Coulomb term, calculate the shift in energy (to first order) of an $l = 0$ hydrogenic state with principal quantum number n, given that the Coulomb wave function at $r = 0$ is

$$\phi_n(0) = \frac{1}{\sqrt{\pi}}\left(\frac{\alpha m}{n}\right)^{\frac{3}{2}} .$$

Give the answer in eV for the $n = 2$ shift.

P3.4 What is the (mass) dimension of a scalar field ϕ in four space-time dimensions? What is the dimension of the coupling constant λ in a '$\lambda\phi^3$' interaction? And of g in a '$g\phi^4$' interaction? What is the dimension of G in a '$G(\bar{\psi}\psi)^3$' interaction?

P3.5 Consider a $\lambda\phi^4$ theory. Given that it is renormalizable, explain why any graph contributing to the process $\phi + \phi \to \phi + \phi + \phi + \phi$ must be finite.

5 GLOBAL AND LOCAL NON-ABELIAN SYMMETRIES

For a much fuller treatment of the material in this section see Chapters 12 and 13 of Volume 2 of the new (third) edition of Aitchison and Hey [2].

Having introduced QED as an example of a *gauge theory* with a *local phase invariance*, we now consider the generalizations of QED which describe the weak and strong interactions between quarks and leptons. These involve a more complicated kind of local phase symmetry, in which the phase factors are (x-dependent) *matrices*, which in general do not commute—that is what 'non-Abelian' means in this context. We shall limit the treatment to the particular ingredients needed for the Standard Model. Note: from now on we shall **omit the hats** on quantum field operators!

5.1 Global non-Abelian symmetry

Consider the Lagrangian for two free fermions of the same mass $m_1 = m_2 = m$

$$\mathcal{L}_2 = \bar{\psi}_1(i\partial\!\!\!/ - m)\psi_1 + \bar{\psi}_2(i\partial\!\!\!/ - m)\psi_2\,; \tag{133}$$

in terms of the 'doublet' field

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \tag{134}$$

it can easily be rewritten as

$$\mathcal{L}_2 = \bar{\psi}(i\partial\!\!\!/ - m)\psi\,. \tag{135}$$

Note that although (135) looks formally like the single-field $\mathcal{L}_{\rm D}$ of (78), it is of course quite different physically, representing two different sorts of particle (e.g. up and down quarks, and their antiparticles). Nevertheless, (135) is invariant under a symmetry rather like (79), namely the 2×2 unitary transformation

$$\psi \to \psi' = U\psi, \quad UU^\dagger = U^\dagger U = 1\,. \tag{136}$$

The U in (136) is a 2×2 matrix of numbers (not field operators) acting on the 2 components of ψ in (134), and they commute with the Dirac γ's. Such unitary 2×2 U's form a group, U(2). Since U in (136) does not involve x, we call (136) a global symmetry. In general, two U's do not commute with each other, and it is called a non-Abelian symmetry.

From elementary properties of determinants we have

$$\det UU^\dagger = \det U.\det U^\dagger = \det U.\det U^* = |\det U|^2 = 1 \tag{137}$$

so that $\det U = e^{-2i\alpha}$, say. We can therefore write

$$U = e^{-i\alpha}\tilde{U} \tag{138}$$

where $\tilde{U}$ has determinant $+1$. Matrices of the form $\tilde{U}$ form the SU(2) group, where the S just means they have unit determinant. The phase factor in (138) corresponds to a simultaneous U(1) transformation of ψ_1 and ψ_2 (with the same phase angle) and leads, as in Section 3.3, to a conservation law of the total number of '1' particles and '2' particles. (For quarks this would be part of baryon number conservation.) The new physics is contained in the $\tilde{U}$ part.

Groups such as SU(2) (and, later, SU(3)) have the important feature that their physically important properties can be found by studying infinitesimal transformations, of the form [cf. (46)]

$$\tilde{U} = 1 - i\xi \tag{139}$$

where ξ is a 2×2 matrix with small entries. The condition $\det U = 1$ gives $\mathrm{Tr}\xi = 0$ (neglecting terms of order ξ^2—see problem P4.1), while $\tilde{U}\tilde{U}^\dagger = 1$ reduces (problem P4.1) to $\xi = \xi^\dagger$. So ξ is a Hermitian traceless matrix. Such a thing depends on only three real parameters (problem P4.1) and can be written as

$$\xi = \boldsymbol{\epsilon}\cdot\boldsymbol{\tau}/2 \tag{140}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)$ are the three parameters, and $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$ are the Pauli matrices (problem P4.2). Thus an infinitesimal SU(2) transformation on the doublet ψ is

$$\psi \to \psi' = (1 - i\boldsymbol{\epsilon}\cdot\boldsymbol{\tau}/2)\psi\,. \tag{141}$$

This should be compared with the infinitesimal version of (69), namely $\psi \to \psi' = (1 - i\epsilon)\psi$, from which it is clear that the 'ϵ' in that case becomes a *matrix* in (141). The form for a finite SU(2) transformation is

$$\psi \to \psi' = e^{-i\boldsymbol{\alpha}\cdot\boldsymbol{\tau}/2}\psi \tag{142}$$

which generalizes (69) (note that for a matrix A, $\exp A = 1 + A + A^2/2! + \cdots$).

Since (141) or (142) are invariances of $\mathcal{L}_2$ we expect an associated conservation law. Indeed, since we have *three* independent transformations (using each of ϵ_i in turn) we expect *three* conservation laws. Following the same steps used in deriving the Noether current for the complex scalar field in §3.2, but this time for the doublet Dirac field ψ, one finds that the three quantities $T_1^\mu(x)$, $T_2^\mu(x)$, $T_3^\mu(x)$ defined by [cf. (70)]

$$T_i^\mu(x) = \bar{\psi}(x)(\tau_i/2)\gamma^\mu\psi(x) \tag{143}$$

satisfy

$$\partial_\mu T_i^\mu(x) = 0 \tag{144}$$

and are therefore symmetry currents. The corresponding 'charges'

$$T_i = \int \psi^\dagger(x)\frac{\tau_i}{2}\psi(x)\,d^3x \tag{145}$$

are conserved. These are the (field theoretic) 'isospin' operators, which have the very interesting property

$$[T_i, T_j] = i\epsilon_{ijk}T_k \tag{146}$$

as can be explicitly checked from (145) (using the proper commutation relations for the ψ fields). A simple example is provided in problem P4.2. The relations (146) are of course exactly the commutation relations of the familiar angular momentum operators, which is why the name iso*spin* was coined; (146) is called the 'SU(2) algebra'. Not coincidentally, the τ's satisfy $[\tau_i/2, \tau_j/2] = i\epsilon_{ijk}\tau_k/2$, the same algebra.

In thinking about more complicated SU(2) multiplets than doublets (which we shall not need to do much) this angular momentum analogy is very helpful. The essential step is to find larger matrices than the 2×2 $\frac{\tau'\mathrm{s}}{2}$, which satisfy commutation relations of the form (146). For example, the three 3×3 matrices t_1, t_2 and t_3, defined by

$$(t_i)_{jk} = -i\epsilon_{ijk}, \tag{147}$$

satisfy $[t_i, t_j] = i\epsilon_{ijk}t_k$ (see problem P4.3). Then if we consider a triplet of three real degenerate fields (bosonic, say)

$$\boldsymbol{\phi} = \begin{pmatrix}\phi_1\\ \phi_2\\ \phi_3\end{pmatrix} \tag{148}$$

with Lagrangian

$$\mathcal{L}_B = \frac{1}{2}\partial_\mu \boldsymbol{\phi}\cdot\partial^\mu\boldsymbol{\phi} - \frac{1}{2}m^2\boldsymbol{\phi}\cdot\boldsymbol{\phi}\,, \tag{149}$$

$\mathcal{L}_B$ is invariant under

$$\boldsymbol{\phi} \to \boldsymbol{\phi}' = (1 - i\boldsymbol{\epsilon}\cdot\boldsymbol{t})\boldsymbol{\phi}\,. \tag{150}$$

Using (147), (150) is equivalent to (problem P4.4)

$$\boldsymbol{\phi}' = \boldsymbol{\phi} + \boldsymbol{\epsilon}\times\boldsymbol{\phi} \tag{151}$$

which should be familiar as the 'infinitesimal rotation' of an ordinary vector.

The SU(2) transformation of (142) can be generalized to the case of *three* degenerate fermion fields. If $\mathcal{L}_3$ is (133) with the addition of $\bar{\psi}_3(i\partial\!\!\!/ - m)\psi_3$, it too can can be written as in (135) where now

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\end{pmatrix}\,. \tag{152}$$

Note particularly that unlike the ϕ's in (148), the ψ's in (152) are complex: each ψ_i contains c_i and $d_i^\dagger$ operators as in (59). $\mathcal{L}_3$ is invariant under $\psi \to \psi' = U\psi$ where U is now an x-independent 3×3 unitary matrix. Extracting the overall phase again, we are left with a global SU(3) transformation. An infinitesimal SU(3) matrix has the form

$$\tilde{U} = 1 - i\chi \tag{153}$$

where χ is a Hermitean traceless 3×3 matrix. Such a χ involves *eight* parameters and can be written as

$$\chi = \boldsymbol{\eta}\cdot\boldsymbol{\lambda}/2 \tag{154}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots \eta_8)$ are the arbitrary parameters and the eight $\boldsymbol{\lambda}$'s are 3×3 Hermitean traceless matrices generalizing the three τ's. They obey the commutation relations

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = if_{abc}\frac{\lambda_c}{2} \tag{155}$$

where the f_{abc} are numbers characteristic of SU(3) (a, b, c all run from 1 to 8). If ψ_1, ψ_2, ψ_3 are taken to be the u, d, s quarks, this global SU(3) symmetry would be the SU(3) of strong interaction flavour symmetry (which however is not exact as m_u, m_d and m_s are not equal). Similarly, if we take 1, 2, 3 to be colour indices we have the exact $\mathrm{SU(3)_c}$ colour symmetry of QCD, which we shall shortly see is a local symmetry. The currents corresponding to the SU(3) symmetry of $\mathcal{L}_3$ are [cf. (143)]

$$G_a^\mu(x) = \bar{\psi}(x)(\lambda_a/2)\gamma^\mu\psi(x) \tag{156}$$

and the associated eight 'charges'

$$G_a = \int \psi^\dagger(x)(\lambda_a/2)\psi(x)\, d^3x \tag{157}$$

generalize the three isospin operators, and obey the commutation relations

$$[G_a, G_b] = if_{abc}G_c \quad , \tag{158}$$

which is called the 'SU(3) algebra'. Note the similarity between (146) and (158).

As in the case of SU(2), larger multiplets are possible too. The key requirement is to find matrices which satisfy (158), since these commutation relations effectively define the group. For SU(3), the only larger multiplet in which we shall be interested is the octet, **8**, which is analogous to the triplet of SU(2).

The matrices for the **8** are defined analogously to the t's of (147), namely $(F_a)_{bc} = -if_{abc}$ where the f's are as in (158). Notice that since there are eight 'charges' G_a, and all the indices a, b, c in (158) run from 1 to 8, the eight matrices F_a are each 8×8. In the same way, the three matrices t_i of (147) are each 3×3, since there are three SU(2) charges. This kind of pattern can be extended to arbitrary SU(N); the 'representation' in which the matrices are equal (with a factor of $-i$) to the 'structure constants' [the ϵ's and f's in (147) and (158)] is generally called the adjoint or regular representation.

5.2 Local non-Abelian SU(2) symmetry

Global symmetries and their associated (possibly approximate) conservation laws are certainly interesting, but they do not have the *dynamical* significance of local symmetries. We saw in Section 3.4 how the 'requirement' of local U(1) symmetry seemed to lead almost automatically to QED, with the symmetry current of the ψ matter fields now playing the role of the dynamical current which, when dotted into the A-field, gives the interaction term in $\mathcal{L}_{QED}$. A similar link between symmetry and dynamics follows if we generalize the preceding non-Abelian global symmetries to local ones. In this section we carry through the analysis for SU(2).

We begin by considering again a fermion doublet as in (135), without yet specifying exactly what the physical application will be. We want to extend the global SU(2) symmetry transformation (142) to the local one

$$\psi(x) \to \psi'(x) = e^{-ig\boldsymbol{\alpha}(x)\cdot\boldsymbol{\tau}/2}\psi(x) \tag{159}$$

by analogy with (79); note that we have slipped in a constant g in the exponent—it will be analogous to the electromagnetic charge e. Clearly, although the $\bar{\psi}m\psi$ part of (135) is still invariant under (159), the $\bar{\psi}i\partial\!\!\!/\psi$ part is not—just as in the U(1) case (80), since the $\partial\!\!\!/$ will pull down a $\partial\!\!\!/\boldsymbol{\alpha}(x)$ factor. As in the U(1) case, we try to compensate this factor by introducing some vector field whose change under an appropriate transformation [accompanying (159)], exactly cancels this $\partial\!\!\!/\boldsymbol{\alpha}(x)$ part. This time, since there are three $\boldsymbol{\alpha}(x)$'s [$\alpha_1(x)$, $\alpha_2(x)$, $\alpha_3(x)$] we immediately see that we need three vector (gauge) fields, called $W_1^\mu(x)$, $W_2^\mu(x)$, $W_3^\mu(x)$, or $\boldsymbol{W}^\mu(x)$ for short.

The key step in constructing the locally U(1) invariant Lagrangian of QED was the replacement of '∂^μ' by '$D^\mu = \partial^\mu + ieA^\mu$' [cf. (81)], together with the transformation '$A^\mu \to A^\mu + \frac{1}{e}\partial^\mu\alpha(x)$' [cf. (83)] for the A-field. Let us have another look at the combination $D^\mu\psi$ in the QED Lagrangian (88). Under the gauge transformation (84),

$$\begin{aligned} D^\mu &= (\partial^\mu + ieA^\mu)\psi \to (\partial^\mu + ieA'^\mu)\psi' \\ &= (\partial^\mu + ieA^\mu + i(\partial^\mu\alpha(x)))e^{-i\alpha(x)}\psi \\ &= [-i(\partial^\mu\alpha(x))e^{-i\alpha(x)}\psi] + e^{-i\alpha(x)}\partial^\mu\psi + ieA^\mu e^{-i\alpha(x)}\psi + [i(\partial^\mu)e^{-i\alpha(x)}\psi] \\ &= e^{-i\alpha(x)}D^\mu\psi \end{aligned} \tag{160}$$

since the bracketed terms cancel. So we have

$$D'^\mu\psi' = e^{-i\alpha(x)}D^\mu\psi \ . \tag{161}$$

In words, this says that the quantity '$D^\mu\psi$' transforms under a *local* U(1) phase transformation just like ψ would under a global one (i.e. it just gets multiplied by a phase factor). So to construct a locally U(1) invariant Lagrangian all we needed to do was multiply $D^\mu\psi$ by $\bar{\psi}$ from the left, since then under the local transformation

$$\bar{\psi}D^\mu\psi \to \bar{\psi}'D'^\mu\psi' = \bar{\psi}e^{i\alpha(x)}\ e^{-i\alpha(x)}D^\mu\psi = \bar{\psi}D^\mu\psi \ , \tag{162}$$

showing that $\bar{\psi}D^\mu\psi$ is indeed locally U(1) invariant. Of course, we also need the γ_μ to get rid of the loose Lorentz index μ, and make $\mathcal{L}$ a Lorentz invariant.

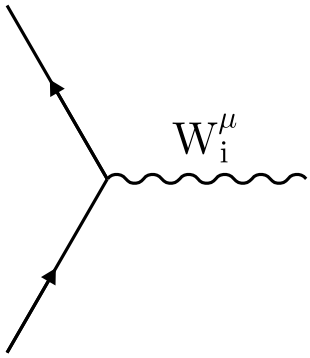


Fig. 11: ψ-ψ-W vertex.

So the key to constructing a locally SU(2) phase-invariant theory is to generalize '$D^\mu\psi$'. The required generalization is

$$D^\mu\psi = (\partial^\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu/2)\psi \tag{163}$$

when acting on an SU(2) doublet field such as ψ. The property required of (163) is that $D^\mu\psi$ should transform under the local symmetry (159) exactly as $\partial^\mu\psi$ does under the global one (142), as we have seen happening in the U(1) case. Then, a term like $\bar{\psi}\not{D}^\mu\psi$ is automatically invariant under local SU(2).

This requirement on $D^\mu\psi$ determines the transformation law of the fields $\boldsymbol{W}^\mu$. The algebra is easier if we consider an infinitesimal transformation

$$\delta\psi = (-ig\boldsymbol{\epsilon}(x)\cdot\boldsymbol{\tau}/2)\psi(x)\,; \tag{164}$$

we then require

$$\delta(D^\mu\psi) = (-ig\boldsymbol{\epsilon}(x)\cdot\boldsymbol{\tau}/2)D^\mu\psi\,. \tag{165}$$

It is a good exercise (problem P4.5) to verify that (165) implies that

$$\delta\boldsymbol{W}^\mu(x) = \partial^\mu\boldsymbol{\epsilon}(x) + g\boldsymbol{\epsilon}(x)\times\boldsymbol{W}^\mu(x)\,, \tag{166}$$

which tells us how the $\boldsymbol{W}^\mu$'s must transform. The first term in (166) is the straightforward analogue of the infinitesimal version of (84), with $\alpha(x) \to e\epsilon(x)$. Comparing the second term of (166) with (151), we see that it implies that the three W-fields form the components of an SU(2) triplet. Thus *the W's carry SU(2) 'charge'*.

We now know the generalization of (135) which makes it locally SU(2) invariant:

$$\mathcal{L}_{2W} = \bar{\psi}(i\not{D} - m)\psi = \bar{\psi}(i\not{\partial} - m)\psi - g\bar{\psi}\gamma_\mu\boldsymbol{\tau}/2\psi\cdot\boldsymbol{W}^\mu, \tag{167}$$

the last term being the generalisation of $\mathcal{L}_{\rm int}$ in QED (equation (82)). We can immediately read off the ψ-ψ-W vertex factor as (Fig. 11)

$$-ig\frac{\tau_i}{2}\gamma^\mu\,. \tag{168}$$

In (168) the index 'i' refers to the SU(2) component of the W field quantum, and 'μ' to the Lorentz component of its polarization vector. Each W-field will have the same kind of mode expansion as the A-field did [Eq. (85)].

We can easily generalize (163) to other SU(2) multiplets than doublets, by using appropriately larger matrices instead of the $\boldsymbol{\tau}/2$. For example, for an SU(2) triplet of fields $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$, (163) becomes

$$D^\mu\phi_i = (\partial^\mu + ig\boldsymbol{t}\cdot\boldsymbol{W}^\mu)\phi_i \tag{169}$$

where the three 3×3 matrices $\boldsymbol{t}$ are defined in (147). Under infinitesimal transformations, this changes by

$$\begin{aligned} \delta(D^\mu \phi_i) &= (-ig\boldsymbol{\epsilon}(x)\cdot\boldsymbol{t})(D^\mu\phi_i) && (170) \\ &= (g\boldsymbol{\epsilon}(x)\times D^\mu\boldsymbol{\phi})_i && (171) \end{aligned}$$

[cf. (150), (151), and (164)].

However, there is still an important part of the non-Abelian analogue of $\mathcal{L}_{\rm QED}$ unaccounted for—namely the bit corresponding to the Maxwell term $-\frac{1}{4}F\cdot F$ for the gauge fields $\boldsymbol{W}^\mu$. Note that, as in the QED case (problem P2.6), a simple mass term involving $\boldsymbol{W}^\mu\cdot\boldsymbol{W}_\mu$ will violate invariance under (166), so these quanta are massless. Clearly we have a problem here in applying this local SU(2)—as we eventually will—to weak interactions, which are very short ranged, and whose quanta are therefore massive. This is where we will need the Higgs mechanism—see Section 6.

To get the non-Abelian '$F\cdot F$' term, the obvious thing might be to consider

$$\partial^\mu A^\nu - \partial^\nu A^\mu \to D^\mu \boldsymbol{W}^\nu - D^\nu \boldsymbol{W}^\mu \tag{172}$$

with D^μ given by (169), since the W's are an SU(2) triplet. The hope would be that by using the D's, $D^\mu\boldsymbol{W}^\nu - D^\nu\boldsymbol{W}^\mu$ would transform under local SU(2) transformations exactly as $\partial^\mu\boldsymbol{W}^\nu - \partial^\nu\boldsymbol{W}^\mu$ does under global ones—i.e. like (171). Then the 'dot product' $(D^\mu\boldsymbol{W}^\nu - D^\nu\boldsymbol{W}^\mu)\cdot(D_\mu\boldsymbol{W}_\nu - D_\nu\boldsymbol{W}_\mu)$ would be a locally invariant '$F\cdot F$' term. Unfortunately it is not quite that simple. The problem is that the W's are a rather special triplet: whereas an ordinary triplet $\boldsymbol{\phi}$ would transform via only the second term in (166), the W's *also* have the first ('non-homogeneous') term as well. You can verify that in fact

$$\delta(D^\mu\boldsymbol{W}^\nu - D^\nu\boldsymbol{W}^\mu) \neq g\boldsymbol{\epsilon}(x)\times(D^\mu\boldsymbol{W}^\nu - D^\nu\boldsymbol{W}^\mu) \tag{173}$$

so that the proposed '$F\cdot F$' term will not work.

With the aid of some hindsight, we can be led to the right answer as follows. Consider, in the U(1) case, the quantity

$$(D^\mu D^\nu - D^\nu D^\mu)\phi \tag{174}$$

where ϕ is any field of charge e and $D^\mu = \partial^\mu + ieA^\mu$. Evaluating (174) one finds (problem P2.5)

$$(D^\mu D^\nu - D^\nu D^\mu)\phi = ieF^{\mu\nu}\phi \tag{175}$$

where $F^{\mu\nu} = \partial^\mu A^\mu - \partial^\nu A^\mu$. This suggests that we should look at the commutator of two covariant derivatives $[D^\mu, D^\nu]$. It does not matter whether we use the D from (163) or (169)—the result is essentially the same for all cases. Using the D^μ from (163) one finds (problem P4.6)

$$[D^\mu, D^\nu] = ig\boldsymbol{\tau}/2\cdot\boldsymbol{F}^{\mu\nu} \tag{176}$$

where

$$\boldsymbol{F}^{\mu\nu} = \partial^\mu\boldsymbol{W}^\nu - \partial^\nu\boldsymbol{W}^\mu - g\boldsymbol{W}^\mu\times\boldsymbol{W}^\nu\,. \tag{177}$$

(Had we used (169) we would have got (176) with $\boldsymbol{\tau}/2 \to \boldsymbol{t}$.) When we now investigate the effect of the local SU(2) transformation (166) on $\boldsymbol{F}^{\mu\nu}$ we find (problem P4.7)

$$\delta\boldsymbol{F}^{\mu\nu}(x) = g\boldsymbol{\epsilon}(x)\times\boldsymbol{F}^{\mu\nu}(x) \tag{178}$$

precisely as desired (but not accomplished) in (173)—i.e. the inhomogeneous part in (166) has been got rid of. Thus $\boldsymbol{F}^{\mu\nu}$ does transform under local SU(2) transformations exactly as if it were an ordinary triplet under global SU(2) transformations and so the quantity

$$\mathcal{L}_W = -\frac{1}{4}\boldsymbol{F}_{\mu\nu}\cdot\boldsymbol{F}^{\mu\nu} \tag{179}$$

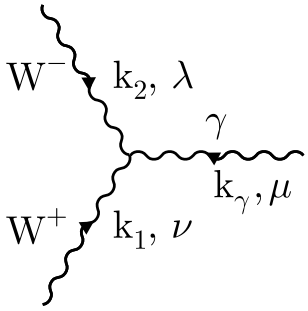


Fig. 12: W-W-γ vertex.

is indeed locally SU(2) invariant. This is the famous *Yang–Mills Lagrangian*, the non-Abelian generalization of the Maxwell Lagrangian. $\boldsymbol{F}^{\mu\nu}$ is the non-Abelian field strength tensor.

The argument leading to (179) has been given in some detail since the result is of fundamental importance. Looking at (177) and (179) it is clear that, unlike the Maxwell term $\mathcal{L}_A$ of (87), the Yang–Mills term $\mathcal{L}_W$ of (179) *includes interactions between the gauge fields*—in addition, of course, to the expected 'free' part

$$-\frac{1}{4}(\partial_\mu \boldsymbol{W}_\nu - \partial_\nu \boldsymbol{W}_\mu)\cdot(\partial^\mu \boldsymbol{W}^\nu - \partial^\nu \boldsymbol{W}^\mu)\,. \tag{180}$$

The free part leads to a W-propagator which is the same as that in rule (v) of section 3.4, with a δ_{ij} factor to 'dot' the W's together. The interactions included in (179) are of two types: W-W-W (trilinear) and W-W-W-W (quadrilinear). This is quite unlike QED, where no fundamental γ-γ vertices are present. It arises here because the W's both 'transmit' the gauge field force and feel it themselves since they are not SU(2) neutral (as the γ was U(1) neutral). Another important point to note is that these self-interactions among the W's come in with a coupling constant which is the same one as appears in the ψ-ψ-W vertex (168)—the W's 'couple universally'.

The physics application of all this is to the SU(2) of the weak interactions (see Section 7). There, the W_1^μ and W_2^μ fields correspond to the charged gauge bosons $W^{\pm\mu}$ (the combination $\frac{1}{\sqrt{2}}(W_1 - iW_2)$ destroys W^+ or creates W^-). As we shall see, the field W_3^μ is a linear combination of the photon γ and Z^0 fields:

$$W_3^\mu = \sin\theta_W A^\mu + \cos\theta_W Z^\mu \tag{181}$$

where θ_W is the 'weak angle', and the SU(2) gauge coupling constant g is related to e by

$$g\sin\theta_W = e\,. \tag{182}$$

We can then pick out the W-W-γ vertex from (179), and find that it is given by

$$ie\,[g_{\nu\lambda}(k_1 - k_2)_\mu + g_{\lambda\mu}(k_2 - k_\gamma)_\nu + g_{\mu\nu}(k_\gamma - k_1)_\lambda] \tag{183}$$

where the momenta and indices are as in Fig. 12. This *unique* electromagnetic coupling of the $W^\pm$ is of precisely the kind needed to make a *renormalizable* (see Section 4) theory of the 'electromagnetic interactions of charged vector bosons'.

5.3 Local SU(3) symmetry: the QCD Lagrangian

Using what has been said about global SU(3) in Section (5.1), and about how to make a global SU(2) symmetry into a local one in Section 5.2, it is straightforward to discuss local SU(3). This is the gauge group of QCD (see the course on QCD), the labels 1, 2, 3 in (152) standing for colour, the ψ's being one flavour of quark. Under a local $\mathrm{SU(3)_c}$ transformation, the triplet (152) transforms by

$$\delta\psi = (-ig_s\boldsymbol{\eta}(x)\cdot\boldsymbol{\lambda}/2)\psi \tag{184}$$

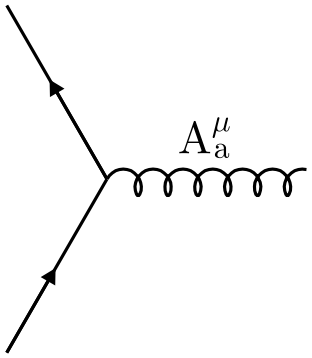


Fig. 13: A-ψ-ψ vertex.

[cf. (154) and (164)], where now there are eight field parameters $\eta_1(x), \eta_2(x) \ldots \eta_8(x)$ going with the eight λ's. To cancel off the unwanted $\partial\boldsymbol{\eta}$ parts which occur when we try to make $\bar{\psi}\partial\psi$ invariant under (184), we now need eight vector gauge fields $A^{\mu}_{a}(x)$, $a = 1, 2, \ldots 8$. These A's transform according to

$$\delta A^{\mu}_{a}(x) = \partial^{\mu}\eta_a(x) + g_s f_{abc}\eta_b(x)A^{\mu}_{c}(x) \tag{185}$$

[cf. (166) and (155)]. The SU(3)$_{\rm c}$ covariant derivative acting on a triplet is

$$D^{\mu}\psi = (\partial^{\mu} + ig_s\boldsymbol{\lambda}/2\cdot\boldsymbol{A}^{\mu})\psi \tag{186}$$

giving the A-ψ-ψ vertex [cf. (168)] of Fig. 13:

$$-ig_s\frac{\lambda_a}{2}\gamma^{\mu}\,. \tag{187}$$

The quanta of the A^{μ}_{a} field are the (eight different) gluons. As in local SU(2), there is an SU(3)$_{\rm c}$ field strength tensor which is [cf. (177)]

$$F^{\mu\nu}_{a} = \partial^{\mu}A^{\nu}_{a} - \partial^{\nu}A^{\mu}_{a} - g_s f_{abc}A^{\mu}_{b}A^{\nu}_{c}\,. \tag{188}$$

The SU(3)$_{\rm c}$ Yang–Mills term is then

$$-\frac{1}{4}F^{\mu\nu}_{a}F_{a\mu\nu} \tag{189}$$

and it contains triple and quadruple gluon couplings, all involving the same 'strong' coupling g_s, and the constants f_{abc} determined from (155). Once again, there is no mass term allowed by invariance under (185), and the gluons are massless. Their propagator is the same as the photon one in rule (v), with a colour factor δ_{ab}.

For one SU(3)$_{\rm c}$ triplet ψ, then, our Lagrangian so far is

$$\mathcal{L} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}F^{\mu\nu}_{a}F_{a\mu\nu} \tag{190}$$

with $D^{\mu}\psi$ given by (186). For many different quark flavours f, the Dirac term is repeated for each, giving

$$\mathcal{L}_{\rm QCD} = \sum_{f}\bar{\psi}_f(i\not{D} - m_f)\psi_f - \frac{1}{4}F^{\mu\nu}_{a}F_{a\mu\nu}\,. \tag{191}$$

Actually, however, matters are not quite that simple. As in QED, we need a gauge-fixing term to produce the gauge field propagator; in the non-Abelian case this turns out to be a more complicated affair, necessitating additional pieces in $\mathcal{L}_{\rm QCD}$ called 'ghost terms'. We shall not give their form here: they are needed only for loop calculations, the details of which we shall not need. The Lagrangian of (191) is adequate at the tree level.

Problems for Lecture 4

P4.1 An 'infinitesimal' SU(2) transformation means one very close to the identity, $\tilde{U} = 1 - i\xi$ where ξ is a matrix whose entries are infinitesimally small. So $\tilde{U} = \begin{pmatrix} 1 - i\xi_{11} & -i\xi_{12} \\ -i\xi_{21} & 1 - i\xi_{22} \end{pmatrix}$. Show that to first order in the ξ's, $\tilde{U}\tilde{U}^\dagger = I$ implies that $\xi = \xi^\dagger$ (i.e. ξ is Hermitean). Also, show (again to first order in the ξ's) that $\det U = 1$ implies $\xi_{11} + \xi_{22} = 0$ (i.e. ξ is traceless). So ξ is a traceless Hermitean matrix, 2×2. Explain why ξ is specified by three real parameters. How many parameters are needed for an infinitesimal SU(N) matrix?

P4.2 The τ-matrices are

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(a) Verify that $[\tau_1/2, \tau_2/2] = i\tau_3/2$. (b) A simple model of the isospin raising operator $\hat{T}_+$ is

$$\hat{T}_+ = (\hat{a}_u^\dagger \; \hat{a}_d^\dagger)(\tau_1/2 + i\tau_2/2) \begin{pmatrix} \hat{a}_u \\ \hat{a}_d \end{pmatrix}$$

where the $\hat{a}^\dagger$'s create u's and d's. Check that $\hat{T}_+ = \hat{a}_u^\dagger \hat{a}_d$ and interpret this. Define also

$$\hat{T}_- = (\hat{a}_u^\dagger \; \hat{a}_d^\dagger)(\tau_1/2 - i\tau_2/2) \begin{pmatrix} \hat{a}_u \\ \hat{a}_d \end{pmatrix}.$$

Show that $\hat{T}_- = \hat{a}_d^\dagger \hat{a}_u$. (c) Evaluate $[\hat{T}_+, \hat{T}_-]$, and check that it is compatible with $[\hat{T}_i, \hat{T}_j] = i\epsilon_{ijk}\hat{T}_k$, where

$$\hat{T}_i = (\hat{a}_u^\dagger \; \hat{a}_d^\dagger)(\tau_i/2) \begin{pmatrix} \hat{a}_u \\ \hat{a}_d \end{pmatrix}.$$

P4.3 The 3×3 matrices $t_1, \; t_2, \; t_3$ are defined by $(t_i)_{jk} = -i\epsilon_{ijk}$ for $i, j, k = 1, 2, 3$ where the index i stands for which t it is, and the j, k indices specify the row and column, respectively, of that ith t matrix. Here ϵ_{ijk} is defined to be 0 if any of i, j, k are equal, +1 if they are a cyclic permutation of '123', and -1 if they are a cyclic permutation of '213'. Write down the 3×3 matrices $t_1, \; t_2, \; t_3$, and verify that $[t_1, t_2] = t_3$.

P4.4 The infinitesimal transformation law of an SU(2) triplet $\boldsymbol{\phi}$ is

$$\begin{pmatrix} \phi_1' \\ \phi_2' \\ \phi_3' \end{pmatrix} = (1 - i\epsilon_1 t_1 - i\epsilon_2 t_2 - i\epsilon_3 t_3) \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}.$$

Calculate the 3×3 transformation matrix explicitly, and show that the transformation can also be written in 'cross product' form $\boldsymbol{\phi}' = \boldsymbol{\phi} + \boldsymbol{\epsilon} \times \boldsymbol{\phi}$.

P4.5 The 'SU(2) covariant derivative' acting on an SU(2) doublet is $D^\mu \psi = (\partial^\mu + ig\boldsymbol{\tau} \cdot \boldsymbol{W}^\mu(x)/2)\psi$. Under an infinitesimal local SU(2) transformation, ψ transforms by

$$\delta\psi = -ig\boldsymbol{\tau} \cdot \boldsymbol{\epsilon}(x)/2 \; \psi \; .$$

The transformation law of $\boldsymbol{W}^\mu$ is determined from the requirement that

$$\delta(D^\mu \psi) = -ig\boldsymbol{\tau} \cdot \boldsymbol{\epsilon}(x)/2 \; (D^\mu \psi) \; .$$

Now the LHS of this equation is

$$\delta[(\partial^\mu + ig\boldsymbol{\tau} \cdot \boldsymbol{W}^\mu(x)/2)\psi]$$

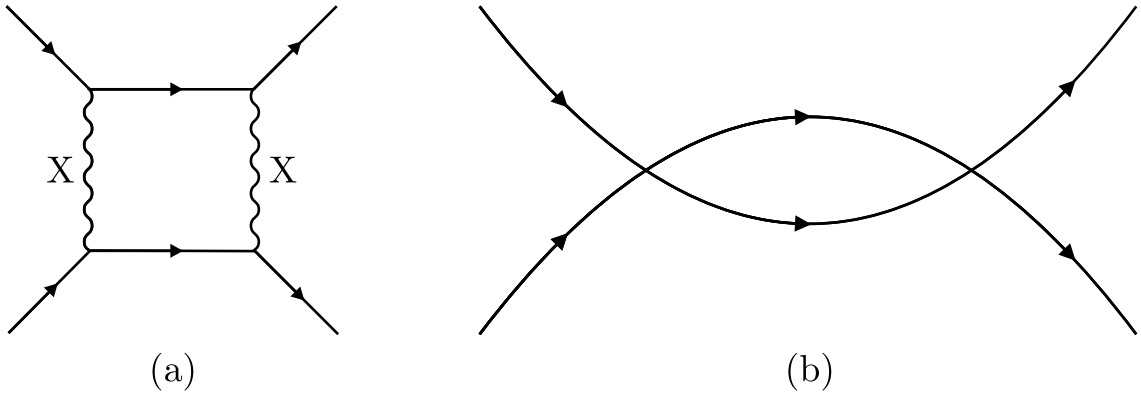


Fig. 14: Two-X exchange in fermion–fermion scattering, and effective four-fermion structure.

$$\begin{aligned} &= ig\boldsymbol{\tau}\cdot(\delta\boldsymbol{W}(x)^\mu/2)\,\psi + (\partial^\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu(x)/2)\delta\psi \\ &= ig\boldsymbol{\tau}\cdot(\delta\boldsymbol{W}(x)^\mu/2)\,\psi + (\partial^\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu(x)/2)(-ig\boldsymbol{\tau}\cdot\boldsymbol{\epsilon}(x)/2)\psi \end{aligned}$$

while the RHS is

$$-ig\boldsymbol{\tau}\cdot\boldsymbol{\epsilon}(x)/2\;(\partial^\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu(x)/2)\psi\;\;.$$

Verify that this implies

$$\delta\boldsymbol{W}^\mu(x) = \partial^\mu\boldsymbol{\epsilon}(x) + g\boldsymbol{\epsilon}(x)\times\boldsymbol{W}^\mu(x)\;\;.$$

P4.6 Check that

$$[\partial^\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu(x)/2, \partial^\nu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}^\nu(x)/2] = ig\boldsymbol{\tau}\cdot\boldsymbol{F}^{\mu\nu}/2$$

where

$$\boldsymbol{F}^{\mu\nu} = \partial^\mu\boldsymbol{W}^\nu(x) - \partial^\nu\boldsymbol{W}^\mu(x) - g\boldsymbol{W}^\mu(x)\times\boldsymbol{W}^\nu(x)\;\;.$$

P4.7 Verify that, under an infinitesimal local SU(2) transformation, $\delta\boldsymbol{F}^{\mu\nu} = g\boldsymbol{\epsilon}(x)\times\boldsymbol{F}^{\mu\nu}$.

6 SPONTANEOUS SYMMETRY BREAKING

See Chapter 21 of Ref. [2].

6.1 Some motivation

In the previous section, an indication was given as to why the relevant theories at current energy scales should be renormalizable theories (a small subclass, incidentally, out of all possible quantum field theories!). We also pointed out how 'universality' phenomena in weak interactions suggested that they are described by a gauge theory, which presumably should be a renormalizable one. On the other hand, we also know that weak interactions are very short-ranged, so their mediating quanta must be massive—and this at once seems to present a barrier to the 'gauge' idea, because (see problem P2.6) a simple gauge boson mass term violates gauge invariance. Perhaps, then, we can have a theory involving massive charged $W^\pm$ bosons, for instance, without it being a gauge theory? Yes, we can, but *it will not be renormalizable*. In fact, the renormalizability of QED has a great deal to do with the gauge symmetry it possesses. Let us try and explain what is wrong with a 'non-gauge theory of massive $W^\pm$'s'.

Consider Fig. 14, which shows some kind of fermion–fermion scattering proceeding, in fourth order of perturbation theory (one loop), via the exchange of two massive vector bosons that we shall call X^μ. To calculate this diagram, we need to know the propagator for X^μ.

For this we need the wave equation for X^μ, which is quite simple to write down. We just replace $\Box$ in the wave equation (91) for A^μ by $\Box + M^2$ where M is the mass of the X^μ:

$$(\Box + M^2)X^\mu - \partial^\mu \partial_\nu X^\nu = 0 \ . \tag{192}$$

To find the propagator, we follow the poor-man's route, putting in a plane wave solution for X^μ, which yields

$$\left[(-q^2 + M^2)\delta^\mu_\nu + q^\mu q_\nu\right] \epsilon^\nu e^{-iq\cdot x} = 0 \ . \tag{193}$$

The propagator should now be proportional to the inverse of the [...] bracket in (193), and (*unlike* the corresponding inverse in (92)!) this does exist and is given by (problem P5.1)

$$\frac{-g^{\mu\nu} + q^\mu q^\nu / M^2}{q^2 - M^2} \ . \tag{194}$$

Note (i) that trouble ensues (the numerator blows up) when $M \to 0$, so already we see that a massless vector particle seems to be a very different kind of thing from a massive one (you cannot just simply take the massless limit); (ii) that if we 'dot' (193) with q_μ we easily deduce $q \cdot \epsilon = 0$ [see below, after (198)].

Now consider the loop integral in Fig. 14. At each vertex we will have a coupling constant factor 'g', which is in fact dimensionless (the interaction will be something like $g\bar{\psi}\gamma_\mu \psi X^\mu$). But, as we warned in Section 4.4, this may not guarantee renormalizabilty, and this is a case where it does not. To get an idea of why not, consider the leading divergent behaviour of Fig. 14. This will be associated with the '$q^\mu q^\nu$' terms in the numerator of (194), so that the leading divergence is effectively

$$\sim \int d^4q \left(\frac{q^\mu q^\nu}{q^2}\right)\left(\frac{q^\mu q^\nu}{q^2}\right)\frac{1}{q}\frac{1}{q} \tag{195}$$

for high q (we are of course not troubling to get all the indices etc. right). But the first two (...)'s in (195) behave like a constant, at large q, so that the asymptotic behaviour is effectively

$$\sim \int d^4q \frac{1}{q}\frac{1}{q} \tag{196}$$

which is *exactly what we would get in a four-fermion theory* !—see Fig. 14, and we know that such a theory is non-renormalizable.

Where have these dangerous powers of q come from? The answer is simple and important. They come from the *longitudinal* polarization state of the massive X particle. We can see this as follows. Consider a free X particle with 4-momentum $q = (q^0, 0, 0, |\boldsymbol{q}|)$, so that the x and y directions are transverse, and the z direction is longitudinal. In the rest frame of the X, the three polarization states can be taken to be

$$\boldsymbol{\epsilon}(\lambda = \pm 1) = \mp 2^{-\frac{1}{2}}(1, \pm i, 0) \ , \quad \boldsymbol{\epsilon}(\lambda = 0) = (0, 0, 1) \ . \tag{197}$$

Boosting to the frame with 4-momentum q, the transverse polarization vectors remain the same, but the longitudinal one becomes

$$\epsilon^\mu(q, \lambda = 0) = M^{-1}(|\boldsymbol{q}|, 0, 0, q^0) \ . \tag{198}$$

Note that $q \cdot \epsilon(q, \lambda = 0) = 0$ is satisfied. At large values of q, $\epsilon^\mu(q, \lambda)$ is therefore proportional to q^μ/M, and this is the origin of such factors in the propagator.

Consider now the photon propagator given by rule (v): there are apparently quite similar factors there too, but they are gauge dependent, and in fact *can be 'gauged away' entirely by choice of* ξ! But, as we have seen, such 'gauging' seems to be possible only in a massless vector theory. A closely related point is that, as we all know, electromagnetic waves are purely transverse: equivalently, free photons exist in only two independent polarization states, instead of the three we might have expected (from the three orientations of their unit spin). The longitudinal state is missing, and it turns out (see Aitchison

and Hey [1] page 188) that this is precisely related to the masslessness of the photon. In the massive X case, all three polarization states are present—and this gives another way of seeing why a massless vector particle is really different from even a very light massive one: there is no smooth naive $M \to 0$ limit.

The above considerations therefore suggest the following line of thought:

– Can we somehow create a gauge theory involving massive vector quanta, such that the offending $q^\mu q^\nu$ bits could be gauged away, making the theory renormalizable? The answer is yes, via the idea of *spontaneous breaking* of the gauge symmetry.

This terminology is contrasted with 'explicit symmetry breaking', in which the observed symmetry breaking is associated with a term in the Lagrangian, in the absence of which the theory would possess some exact symmetry. For example, to the extent that the up and down quark masses are equal, we have approximate SU(2) flavour symmetry of the QCD Lagrangian. But it is also possible to have a symmetrical Lagrangian, while the particle states and other physical observables seem to show no obvious (even approximate) sign of the symmetry. This is the 'spontaneously broken' case. This language is borrowed from condensed matter physics, where the ferromagnet is the frequently quoted example. The (Heisenberg) Hamiltonian is certainly rotationally invariant, yet below the transition temperature the spins are thought of as lining up in some particular direction, breaking the rotational symmetry 'spontaneously'.

In the case of a field theory, there are striking differences in the physical consequences depending on whether the symmetry that is spontaneously broken is a global or a local one. In the global case, a general result due to Goldstone [3] and others states that spontaneous breaking of a continuous symmetry is always associated with the appearance of a massless particle, or particles, called 'Goldstone bosons'. In the local case, these Goldstone bosons become the longitudinal components of the gauge field(s)—which, before symmetry breaking, always had only the two transverse components. The total of three 'spin' components in all is exactly what is required for a *massive* vector field. This is the essence of the theoretical loophole which allows gauge bosons to be massive even though the Lagrangian is locally (gauge-) invariant (cf. problem P2.6), and which is invoked to give masses to the W and the Z bosons in the Standard Model.

We begin with the simpler case of spontaneously broken global symmetry, which is of physical importance in its own right in the non-Abelian case (Section 6.3).

6.2 Spontaneously broken global U(1) symmetry

See Chapter 17 of Ref. [2].

We consider a simple classical field theory which shows the effect we want to study. Let ϕ be a complex scalar field, described by the Lagrangian

$$\mathcal{L}_\phi = \partial_\mu \phi^* \partial^\mu \phi - V(\phi) \tag{199}$$

where the potential is taken to have the form ($\lambda > 0$)

$$V(\phi) = -\mu^2 \phi^* \phi + \frac{\lambda}{4}(\phi^* \phi)^2 \,. \tag{200}$$

Clearly $\mathcal{L}_\phi$ is invariant under the global U(1) symmetry

$$\phi \to \phi' = e^{-i\alpha}\phi \,. \tag{201}$$

(Note that a term like $(\phi^*\phi)^3$ would also be invariant under (201), but this would be a non-renormalizable interaction in the quantum theory of $\mathcal{L}_\phi$, so we exclude it.)

Application of the Euler–Lagrange equation yields the equation of motion

$$(\Box - \mu^2)\phi = -\frac{\lambda}{2}|\phi|^2 \phi \,. \tag{202}$$

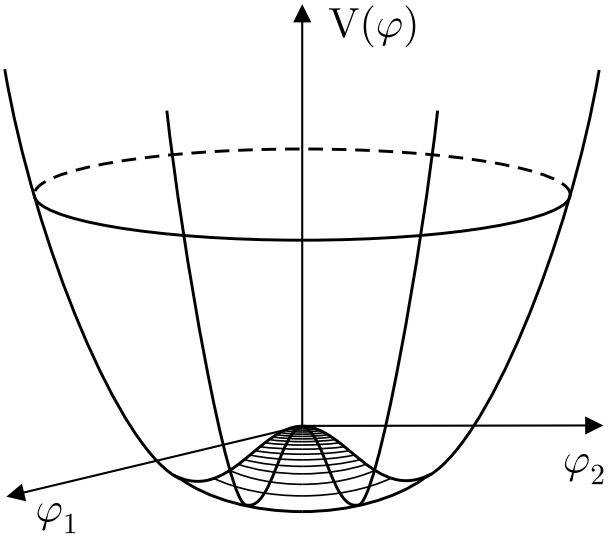


Fig. 15: The 'wine-bottle' potential of (200).

This is nearly the standard Klein–Gordon equation for ϕ (with an interaction term on the right-hand side)—except for the fact that '$-\mu^2$' has the wrong sign for a mass term! This prevents us from making any quantum interpretation of (199) as yet; we therefore concentrate on $V(\phi)$ regarded simply as the potential energy of the classical field.

As a first step to understanding (199), we try to identify the configuration(s) of minimum energy, about which the system might be expected to oscillate. Generally, the energy will be a minimum when ϕ is a constant, which reduces the kinetic terms to zero. The minimum energy is then reached at the minimum of $V(\phi)$. This occurs at

$$|\phi| = v/\sqrt{2}, \quad v = 2\mu/\lambda^{1/2}\,, \tag{203}$$

where v is referred to as the 'symmetry breaking parameter'. To have a clearer picture, it is helpful to introduce two real fields ϕ_1 and ϕ_2 by

$$\phi = (\phi_1 - i\phi_2)/\sqrt{2} \tag{204}$$

and also the 'polar' variables

$$\phi = (\rho/\sqrt{2})e^{i\theta/v}\,, \tag{205}$$

where the v is inserted so that θ has the same dimensions as ρ. Figure 15 shows $V(\phi)$ versus ϕ_1 and ϕ_2, from which it is obvious that the minimum of V is not at $\phi_1 = \phi_2 = 0$. In fact, there is *no* unique minimum point—rather, any value on the circle $\phi_1^2 + \phi_2^2 = v^2$ or equivalently $\rho = v$ will do. Before proceeding further, we briefly outline the condensed matter analogue of (199) and (200) which we mentioned earlier—namely the ferromagnet. In this case, one considers the free energy as a function of the magnetization $\boldsymbol{M}$ at a given temperature T, and makes an expansion of the form

$$F \approx F_0(T) + \frac{1}{2}\mu^2(T)\boldsymbol{M}^2 + \frac{1}{4}\lambda(T)(\boldsymbol{M}^2)^2 + \cdots\,, \tag{206}$$

valid for small magnetization. If the parameter μ^2 is positive, it is easy to see that F has a simple 'bowl' shape as a function of $|\boldsymbol{M}|$, with a minimum at $|\boldsymbol{M}| = 0$. This is the case for T greater than the ferromagnetic transition temperature $T_{\rm C}$. However, if one assumes that $\mu^2(T)$ becomes negative for $T < T_{\rm C}$ (so that $\mu^2(T_{\rm C}) = 0$), then F will now look like Fig. 15 and the minimum free energy will occur for $|\boldsymbol{M}| \neq 0$. The interpretation is that in this case the ground state will be magnetized. Any direction

of $\boldsymbol{M}$ is possible (only $|\boldsymbol{M}|$ is specified); but when the system does settle into one actual configuration with $\boldsymbol{M} \neq \mathbf{0}$ the original full rotational invariance of (206) is lost—the magnetization, and the breaking of the symmetry, has occurred 'spontaneously'.

In the same way, any particular minimum on the circle $\rho = v$ will select out a particular θ in (205), breaking 'spontaneously' the invariance (201).

In quantum field theory, particles are thought of as excitations from a ground state, which we call 'the vacuum'. Figure 15 strongly suggests that if we want a decent quantum interpretation of (199), we should consider expanding the fields about a point on the circle of minima, about which stable oscillations are likely. Any such point represents a possible vacuum state in which

$$\langle 0|\, \phi_1^2 + \phi_2^2 \,|0\rangle = v^2\,, \qquad \text{or } \langle 0|\, \rho \,|0\rangle = v\,. \tag{207}$$

Bearing in mind [cf. (200)] that for a field with a conventional (positive) mass^2 parameter the potential would be U-shaped, we might guess that 'radial' oscillations in Fig. 15 would correspond to a conventional massive field, while 'angle' oscillations—which pass through all the degenerate minima (vacua)—have no 'restoring force' and are massless. Accordingly, we set [cf. (205)]

$$\phi(x) = \frac{1}{\sqrt{2}}(v + h(x))e^{-i\theta(x)/v} \tag{208}$$

and find that $\mathcal{L}_\phi$ becomes (problem P5.2)

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu h \partial^\mu h - \mu^2 h^2 + \frac{1}{2}\partial_\mu\theta\partial^\mu\theta + \frac{\mu^4}{\lambda} + \text{terms cubic and quartic in } \theta, h\,. \tag{209}$$

Equation (209) exhibits the desired form of a conventional scalar field h with mass $\sqrt{2}\mu$ and a massless field θ, together with interaction terms. In particular, the quantum version of (209) will have $\langle 0|h(x)|0\rangle = \langle 0|\theta(x)|0\rangle = 0$, consistent with (207), so that h and θ will have the usual mode expansions (of the form (19) for example), allowing the usual particle interpretation. (The constant term in (209), which does not affect equations of motion, reflects the fact that $V(\min) = -\mu^4/\lambda$.) Note that the symmetry (201), which is evident in (199), is well and truly *hidden* in (209)!

This model (due originally to Goldstone [3]) contains the essence of spontaneous symmetry breaking in field theory: a non-zero value of a field in the ground state (vacuum), a zero mass mode or modes (the Goldstone bosons), and a massive excitation or excitations in the directions 'perpendicular' to the degenerate ground states.

It is interesting to find out what happens to the symmetry current corresponding to the invariance (201). Following the usual procedure, this current is

$$j^\mu_\phi \;=\; i\left\{\phi^\dagger\partial^\mu\phi - (\partial^\mu\phi)^\dagger\phi\right\} \;=\; v\partial^\mu\theta + 2h\partial^\mu\theta + h^2\partial^\mu\theta/v\,. \tag{210}$$

The presence of the term involving just the *single* field θ is very remarkable: it tells us that (in the quantum theory) there is a non-zero matrix element of the form

$$\langle 0|\, j^\mu_\phi(0)\,|\theta\rangle = -ip^\mu v\,, \tag{211}$$

where $|\theta\rangle$ stands for a state with one Goldstone boson θ, with momentum p^μ. That is, the symmetry current connects the Goldstone boson to the vacuum, with an amplitude proportional to the symmetry breaking parameter. In the case of spontaneously broken chiral SU(2)$_{\text{f5}}$ symmetry (Section 6.3 below), the analogue of j^μ_ϕ is the current of the global axial SU(2) symmetry A^μ_i, and there are three θ modes which are identified with the physical pions. The parameter v in the corresponding equation (211) is then f_π (~ 94MeV), the constant which enters into the pion decay $\pi \to \ell\nu$.

Although by the ansatz (208) we seem to have arrived at a viable particle interpretation of (199), we might well ask: how would such a negative (mass)^2 term arise in quantum field theory? One possible answer is that, as with the ferromagnetic analogy, the coefficient μ^2 in (200) could be temperature dependent: perhaps at extremely high temperatures, such as prevailed in the early universe, μ^2 had the opposite sign, corresponding to a conventional mass term. In that case the potential would have a simple minimum at the origin, and the symmetry would not be spontaneously broken until T dropped below some $T_{\rm C}$, where $\mu^2(T_{\rm C}) = 0$. This simple picture is indeed popular in models of the early universe, where such phase transitions are proposed. On the other hand, it may be that some theory might predict the coefficient μ^2 in (200) to be negative, in a particular case. Or, one might simply postulate a $V(\phi)$ of the form (200), so as to 'trigger' the desired breakdown. The last alternative is essentially what is done in the Higgs sector of the Standard Model—as we shall discuss in Section 6.5 and Section 7.

6.3 Spontaneously broken global chiral symmetry

See Section 12.3.2 and Chapter 17, of Ref. [2].

The Dirac Lagrangian for a single massless fermion,

$$\bar{\psi} i \not{\partial} \psi \tag{212}$$

is invariant not only under the ordinary global U(1) symmetry of (69), but also under the 'γ_5-version' of it, namely

$$\psi \to \psi' = e^{-i\eta\gamma_5}\psi \ . \tag{213}$$

This can be easily verified directly, using

$$\gamma^0\gamma^5 = -\gamma^5\gamma^0, \quad \gamma^i\gamma^5 = -\gamma^5\gamma^i \ , \tag{214}$$

but it will be useful later to expand the discussion now to cover this type of symmetry, not considered previously. We may write

$$\psi = \frac{(1-\gamma_5)}{2}\psi + \frac{(1+\gamma_5)}{2}\psi \ \equiv \ \psi_{\rm L} + \psi_{\rm R} \ . \tag{215}$$

The ordinary (infinitesimal) U(1) symmetry (69) is then

$$\delta\psi_{\rm R} = -i\epsilon\psi_{\rm R}, \quad \delta\psi_{\rm L} = -i\epsilon\psi_{\rm L} \tag{216}$$

while the infinitesimal version of (213) is

$$\delta\psi_{\rm R} = -i\eta\psi_{\rm R} \, , \quad \delta\psi_{\rm L} = +i\eta\psi_{\rm L} \, . \tag{217}$$

Transformations such as (217), which act differently on the L and R components are called 'chiral'. Using (214), (215) can be written as

$$\bar{\psi} i \not{\partial} \psi = \bar{\psi}_{\rm L} i \not{\partial} \psi_{\rm L} + \bar{\psi}_{\rm R} i \not{\partial} \psi_{\rm R} \, , \tag{218}$$

which clearly exhibits both the symmetries (216) and (217). It is also manifestly L $\leftrightarrow$ R symmetric, which means it conserves parity. On the other hand, a mass term $m\bar{\psi}\psi$ becomes

$$m\bar{\psi}\psi = m(\bar{\psi}_{\rm L}\psi_{\rm R} + \bar{\psi}_{\rm R}\psi_{\rm L}) \tag{219}$$

which is invariant under (216) but not under (217), while still preserving parity.

Consider then $\mathcal{L}_{\rm QCD}$ of (191), in the limit in which some quark masses—in particular the lightest, m_u and m_d—are regarded as negligible. The fact that $\not{\partial}$ in (212) is replaced by $\not{D}$ clearly makes no

difference to the preceding discussion, which depended only on (214). Thus in this limit $\mathcal{L}_{\rm QCD}$ will be invariant under the γ_5-version of (141), namely

$$\delta\psi = -i\boldsymbol{\eta}\cdot\boldsymbol{\tau}/2\,\gamma_5\,\psi\,, \tag{220}$$

which is a chiral 'SU(2)$_{\rm f5}$' transformation. Now this cannot be realized as an exact symmetry in nature, or else for every non-strange baryon made of u and d quarks there would have to exist another one, degenerate in mass, but with the opposite parity. The reason is worth pausing over.

Associated with the invariance (220) will be three conserved charges, just as in (141)–(145), namely

$$T_i^5 = \int \psi^\dagger(x)\frac{\tau^i}{2}\gamma_5\psi(x)d^3x\ \ . \tag{221}$$

In this case, however, these objects are 'pseudoscalars' (because of the γ_5)—meaning that they will change the parity of any state they act on. Thus whereas the ordinary isospin raising operator $T_+ = T_1 + iT_2$ has the action $T_+|d\rangle = |u\rangle$, where u and d are degenerate in mass because $[T_+, H] = 0$ (T_+ is a constant of the motion), in the case of T_+^5 we must have

$$T_+^5|d\rangle = |\tilde{u}\rangle \tag{222}$$

where $\tilde{u}$ is an 'up' state, degenerate in mass with $|u\rangle$ (because $[T_+^5, H] = 0$ also), but with opposite parity.

Such negative parity analogues of all non-strange baryons are not seen experimentally. One might of course blame this on the finite mass of the u and d quarks, but this is implausible. Instead, we try the idea that this chiral symmetry is spontaneously broken. In that case, we expect three massless Goldstone bosons (corresponding to the three independent SU(2) chiral transformations), and we can interpret $\tilde{u}\rangle$ of (222) as being really $|u +$ massless pseudoscalar boson$\rangle$, thus producing a state degenerate with u in mass, but of opposite parity! These three massless Goldstone bosons are identified with the *pions*—thereby explaining their anomalously low mass (by comparison with that of the ρ-meson, for example). The mass of the physical pion is not, of course, strictly zero, and this is attributed to small non-zero quark masses in the original QCD Lagrangian. Still useful, though more 'explicitly' broken than this chiral SU(2), is the chiral flavour SU(3) analogue, in which we suppose $m_s \approx 0$—the Goldstone bosons are then the kaons.

Remarkably enough, these ideas are also relevant to the weak interactions. In this case, as we shall see, the interaction is most definitely not left–right symmetric (it violates parity)—indeed the 'V−A' structure means that the weak gauge fields couple only to the $\psi_{\rm L}$ components of the fermions, and not to the $\psi_{\rm R}$ components at all. This means that the corresponding local gauge symmetry is of the form

$$\delta\psi_{\rm L} = -i\boldsymbol{\epsilon}\cdot\boldsymbol{\tau}(x)/2\,\psi_{\rm L} \tag{223}$$

$$\delta\psi_{\rm R} = 0\,, \tag{224}$$

for a 'weak doublet' such as

$$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}. \tag{225}$$

But this implies that any mass term of the form (219), which treats $\psi_{\rm L}$ and $\psi_{\rm R}$ the same, will break this 'left-handed' gauge symmetry. Although the neutrinos were usually taken to be massless, the other leptons are definitely not, nor are the quarks. Thus, curiously enough, there is another 'mass problem' with the weak interactions: they would like not only the W and Z bosons but also the fermions to be massless. Once again, we shall have to suppose that the fermion masses arise 'spontaneously', if we want to save the (weak) gauge symmetry. In the Standard Model, one appeals to the same mechanism (the Higgs field) to give mass to the gauge bosons and to the fermions, which is an economical but not necessary step; see Section 7.

It is now time to turn to spontaneously broken local symmetries, concentrating on those relevant to the Standard Model.

6.4 Spontaneously broken local U(1) symmetry: the Abelian Higgs model

See Section 19.3 of Ref. [2].

The U(1) Higgs model is just $\mathcal{L}_\phi$ of (199) extended so as to be locally U(1) invariant; it provides a beautifully simple model for investigating what happens when a *gauge* symmetry is spontaneously broken. To make (199) locally U(1) invariant, we need only replace ∂'s by D's as in (81), and add the Maxwell piece, giving

$$\mathcal{L}_h = [(\partial_\mu + ieA_\mu)\phi]^\dagger\,[(\partial^\mu + ieA^\mu)\phi] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(\phi) \tag{226}$$

where V is still (200), and of course $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Equation (226) is invariant under the local version of (201), namely

$$\phi \to \phi'(x) = e^{-i\alpha(x)}\phi(x) \tag{227}$$

when accompanied by a gauge transformation on A^μ

$$A^\mu \to A^{\mu\prime} = A^\mu + \frac{1}{e}\partial^\mu\alpha \tag{228}$$

as in Section 3.4. Before proceeding further, we note at this stage that we have four field degrees of freedom—two in ϕ and two in the massless A^μ ($F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$).

Now we have learned that the form of V in (200) does not lend itself to a natural particle interpretation, which only appears after making the 'shift to the minimum', as in (208). But there is a remarkable difference between the local and global cases. In the local case, the phase of ϕ is completely arbitrary, since any change in $\theta(x)$ in (208) can be compensated by an appropriate transformation (228) on A^μ, leaving $\mathcal{L}_h$ the same as before. Thus in fact the 'θ' field in (208) can be 'gauged away' altogether, if we like! This must mean that the massless Goldstone boson, described precisely by θ in the quantum theory, somehow no longer appears. This is the first unexpected result in the local case (and it reminds us of our desire to 'gauge away' those longitudinal polarization states ...).

However, we cannot simply 'lose' degrees of freedom. Somehow the system must keep track of the fact that we started with four. To see what has happened, we substitute (208) into (226) with $\theta = 0$; i.e. set

$$\phi = \frac{1}{\sqrt{2}}(v + h(x)) \tag{229}$$

in $\mathcal{L}_h$. We find then (problem P5.3)

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h\,\partial^\mu h - \mu^2 h^2 + \frac{\mu^4}{\lambda} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu + \text{ interaction terms }, \tag{230}$$

where A^μ has to be understood as the gauge field after the transformation needed to reduce ϕ to (229). Equation (230) shows the second 'Higgs miracle': we see that the A^μ field now has a *mass*, equal to ev where v is the symmetry breaking parameter. The missing degree of freedom has reappeared as the third (longitudinal) polarization state of the massive field A^μ. The fourth degree of freedom is still there, the massive h field as in (209).

Can such miracles ever occur? The answer is undoubtedly yes, at least in the non-relativistic case. The low-energy version of $\mathcal{L}_h$ is just the Ginzburg–Landau (GL) approximation for (again) the free energy in a superconductor. In this case (see Section 19.2 of Aitchison and Hey [2] for example) ϕ represents a composite (rather than elementary) field, such that $|\phi|^2$ is the density of bound Cooper pairs (of e^-e^-). Also, the mass for the A field implies that the field is exponentially attenuated inside the superconductor, with a penetration length of order $1/ev$; this is the Meissner effect. It is worth noting that the GL free energy is not to be regarded as a fundamental theory, which must of course be derived from the physical electron–electron and electron–lattice interactions; this is what the BCS theory is all about,

and the GL free energy is a phenomenological expression embodying much of the important physics of the BCS theory. In particle physics the question of whether the ϕ field in the Standard Model (see Section 7) is elementary or composite is completely unknown. However, whatever the truth of that may be, it seems pretty well inevitable that some such field, or effective field, is required to give mass to the W and Z (see Section 6.5, and Section 7)—and in that case it should have its own excitation quantum, the *Higgs boson*: hence the intense interest in hunting for it!

Before proceeding further we can at this stage read off from (230) the propagator for the massive vector A-field. As in the discussion following (193), we need to invert the quantity $P_{\mu\nu}(M_A) = [(-k^2 + M_A^2)g_{\mu\nu} + k_\mu k_\nu]$, where $M_A = ev$ here. As we saw, this does have a straightforward inverse, leading to the propagator

$$i\frac{(-g_{\mu\nu} + \frac{k_\mu k_\nu}{M_A^2})}{k^2 - M_A^2}\,. \tag{231}$$

We see that (231) makes no sense as $M_A \to 0$, reflecting the difficulty with the massless limit of the massive theory. A more technical point concerns the fact that (231) obtains only when the special choice of gauge, $\theta = 0$, is made as in (229). In general, the vector propagator will contain a gauge parameter ξ like the massless propagator of rule (v): this is after all a gauge theory! Rule (v) becomes

• rule (v)′ a factor $i[-g^{\mu\nu} + \frac{(1-\xi)q^\mu q^\nu}{q^2 - \xi M^2}]/(q^2 - M^2)$ for an internal massive gauge boson carrying 4-momentum q, where ξ is a gauge parameter ($\xi \to \infty$ gives the 'naive' vector boson propagator).

Note that for finite ξ, this propagator has a large q behaviour $\sim 1/q^2$, which is good enough to make Fig. 14 convergent! This, then, is the essential clue as to how we can have a renormalizable theory with massive gauge bosons. The gauge $\xi \to \infty$ is called 'unitary gauge': in this gauge there is no visible sign of the scalar ϕ-field. But note that in gauges with ξ finite, the scalar field will also be present with a ξ-dependent propagator (associated with the degree of freedom suppressed in (229)); the complete theory is nevertheless always ξ-independent. Further discussion of this is contained in Section 19.5 of Aitchison and Hey [2] for example.

Returning to (226), we can again look at the electromagnetic current in this 'spontaneously broken local U(1)' model. The gauge invariant form of (210) is

$$\begin{aligned} j^\mu_{\text{e.m.}} &= ie\left[\phi^\dagger(\partial^\mu + ieA^\mu)\phi - \text{ complex conjugate}\right] \\ &= ie(\phi^\dagger\partial^\mu\phi - (\partial^\mu\phi^\dagger)\phi) - 2e^2A^\mu\phi^*\phi\,. \end{aligned} \tag{232}$$

Inserting (208) into (232) (this time in a gauge such that $\theta \neq 0$) we find [cf. (210)]

$$j^\mu_{\text{e.m.}} = -e^2v^2A^\mu + ev\partial^\mu\theta + \text{interaction terms}\,. \tag{233}$$

Equation (233) tells us that there is a 'screening current' (the first term on the RHS) which leads to a mass ev of the A-field, once again; the second term shows that—as in (211)—the vacuum couples to the 'would-be Goldstone boson' (which has become the longitudinal part of the A-field) via the electromagnetic current.

This is an important observation as it leads to a somewhat different way of understanding the 'mechanism' whereby a gauge particle can become massive. In Section 5.1 we introduced the photon self-energy $\Pi_{\rho\sigma}$ which had the general form

$$\Pi_{\rho\sigma} = (g_{\rho\sigma}q^2 - q_\rho q_\sigma)\Pi^\gamma(q^2)\,. \tag{234}$$

When all the self-energy insertions are summed up, and after renormalization, the photon propagator has the form [cf. (111)]

$$-ig^{\mu\nu}/q^2\left(1 - \bar{\Pi}_\gamma(q^2)\right)\,, \tag{235}$$

Fig. 16: Massless Goldstone boson coupling to photon.

in the Feynman gauge. The existence of the matrix element

$$\langle 0|j^{\mu}_{\rm e.m.}(0)|\theta\rangle = -iq^{\mu}ev \tag{236}$$

means that $\Pi_{\rho\sigma}$ will now receive a contribution from the diagram of Fig. 16, where the dotted line represents the massless θ quantum. This is now a tree diagram, not a loop as in the e^+e^- contribution of Fig. 9(a), and so the contribution to $\Pi_{\rho\sigma}$ will involve simply the (massless) θ-propagator, with no momentum integration. The γ-θ vertex is given by (236), with the result that the contribution to $\bar{\Pi}_{\gamma}(q^2)$ in (235) is

$$\bar{\Pi}^{\theta}_{\gamma}(q^2) = e^2v^2/q^2\,, \tag{237}$$

so that the pole in the photon propagator (235) is now at $q^2 = e^2v^2$, and the photon has a mass ev, as before. We have been casual about questions of gauge choice in this argument, but the essential point is valid: a gauge quantum can acquire mass if (for some reason) its vacuum polarization function has a zero mass pole (see the *Discussion point* after (99)). This pole can be associated with the 'elementary' massless quantum in a Higgs potential of the form (200), but it does not have to be. The massless quantum could equally well be a bound state in some strongly-interacting fermion–antifermion channel—in particular, a Goldstone boson arising from the spontaneous breaking of some global symmetry in a purely fermionic theory, for instance. All that is necessary is that it has a coupling of the form (236). The point of this latter interpretation is that only the product 'ev' has significance—there is no sign of Fig. 15, or of 'v' alone as the vacuum value of a scalar field. Theories of this latter type do seem to produce a natural 'dynamical' mechanism for gauge boson mass generation. Both the '$t\bar{t}$' models (Nambu [4]; Miransky et al. [5], [6]; Bardeen et al. [7]), and technicolour (Farhi and Susskind [8]), are of this type, but neither seem to be favoured by experiment. In the electroweak theory it is of course the W and Z particles that we want to be massive (while still being gauge bosons), not the photon. We therefore need to extend the above to the (non-Abelian) SU(2) case.

6.5 Spontaneously broken SU(2)×U(1) symmetry: the gauge and Higgs field sectors of the electroweak theory

See Section 19.6 of Ref. [2].

We shall confine ourselves to the particular case which we need for the electroweak theory. We consider a complex scalar (spin-0) SU(2) doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{238}$$

where the complex ϕ^+ field destroys positively charged particles and creates negatively charged ones, and the complex ϕ^0 field creates neutral particles and antiparticles (a hadronic analogy would be the K^+ and K^0 fields under hadronic SU(2)$_{\rm f}$). The Lagrangian

$$\mathcal{L}_{\Phi} = (\partial_{\mu}\phi)^{\dagger}(\partial^{\mu}\phi) + \mu^2\phi^{\dagger}\phi - \frac{\lambda}{4}(\phi^{\dagger}\phi)^2 \tag{239}$$

then exhibits a global SU(2) invariance of the form [cf. (159)]

$$\phi \to \phi' = \exp(-i\boldsymbol{\alpha}\cdot\boldsymbol{\tau}/2)\phi\,, \tag{240}$$

but this is spontaneously broken, the minimum of the potential in (239) occurring at [cf. (207)]

$$(\phi^\dagger\phi)_{\min} = 2\mu^2/\lambda \equiv v^2/2\,. \tag{241}$$

As in the U(1) case, we interpret (241) in the quantum theory as [cf. (207)]

$$\langle 0|\,\phi^\dagger\phi\,|0\rangle = v^2/2\,, \tag{242}$$

so that the ϕ-field has a non-zero value in the vacuum. Once again, we exclude higher powers of $\phi^\dagger\phi$ in (237) on grounds of renormalizability.

As before, in order to get a sensible particle spectrum we must 'shift' the fields so as to deal with stable oscillations about the minimum (vacuum) given by (242). So we need to define '$\langle 0|\phi|0\rangle$' and expand about it, as in (207) and (208). In the present case, however, the situation is more complicated than (208), since the complex doublet (238) contains four real fields, parametrized for example as

$$\phi^+ = \frac{1}{\sqrt{2}}(\phi_1 - i\phi_2)\,, \quad \phi^0 = \frac{1}{\sqrt{2}}(\phi_3 - i\phi_4)\,; \tag{243}$$

(242) then becomes

$$\langle 0|\,\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\,|0\rangle = v^2\,. \tag{244}$$

It is evident that we have a lot of freedom in choosing the $\langle 0|\phi_i|0\rangle$ so that (244) holds, and it is not at first obvious what an appropriate generalization of (207) and (208) might be.

Furthermore, in this more complicated (non-Abelian) situation a qualitatively new feature can arise: it may happen that the chosen condition $\langle 0|\phi_i|0\rangle \neq 0$ is *invariant* under some subset of the allowed symmetry transformations. This would effectively mean that this particular choice of the vacuum state respected that subset of symmetries, which would therefore not be 'spontaneously broken' after all. Since each broken symmetry is associated with a massless Goldstone boson, we would then get fewer of these bosons than expected.

Just this happens (by design!) in the present case. To understand how it works, we must first recognize that, in addition to the global SU(2) symmetry of (4.41), $\mathcal{L}_\Phi$ of (240) is also invariant under a completely independent global U(1) symmetry of the form

$$\phi \to \phi' = e^{-i\beta}\phi\ , \tag{245}$$

which just means that the phases of the upper and lower components of ϕ in (238) change simultaneously by the same amount. Thus the full symmetry of (239) is global SU(2)×U(1) (which will be made local in a moment, as is required in the Standard Model).

Suppose then that we could choose the $\langle 0|\phi_i|0\rangle$ so as to break this SU(2)×U(1) symmetry completely: we would then expect four massless fields. Actually, however, it is not possible to make such a choice. An analogy may make this point clearer. Suppose we were considering just SU(2), and the field ϕ was an SU(2)-triplet. Then we could always write $\langle 0|\,\boldsymbol{\phi}\,|0\rangle = v\boldsymbol{n}$ where $\boldsymbol{n}$ is a unit vector; but this form is invariant under rotations about the $\boldsymbol{n}$-axis, irrespective of where that points. In the present case, by using the freedom of global SU(2)×U(1) phase changes, an arbitrary $\langle 0|\phi|0\rangle$ can be brought to the form

$$\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}. \tag{246}$$

In considering what symmetries are respected or broken by (246), it is easiest to look at infinitesimal transformations. It is then clear that the particular transformation

$$\delta\phi = -i\epsilon(1+\tau_3)\phi \tag{247}$$

(which is a combination of (245) and the 'third component' of (240)) is still a symmetry of (246) since

$$(1+\tau_3)\begin{pmatrix}0\\ v/\sqrt{2}\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}, \tag{248}$$

so that

$$\langle 0|\phi|0\rangle = \langle 0|\phi+\delta\phi|0\rangle \ ; \tag{249}$$

we say that 'the vacuum is invariant under (247)', and when we look at the spectrum of oscillations about that vacuum we expect to find only three massless bosons, not four.

Oscillations about (246) are conveniently parametrized by

$$\phi=\exp(-i(\boldsymbol{\theta}(x)\cdot\boldsymbol{\tau}/2)v)\begin{pmatrix}0\\ \frac{1}{\sqrt{2}}(v+H(x))\end{pmatrix}, \tag{250}$$

which is to be compared with (208). Inserting (250) into (239) (see problem P5.4), we easily find that no mass term is generated for the $\boldsymbol{\theta}$ fields, while the H field piece is

$$\mathcal{L}_H=\frac{1}{2}\partial_\mu H\partial^\mu H-\mu^2H^2+\text{ interactions} \tag{251}$$

just as in (209), showing that $m_H=\sqrt{2}\mu$.

As noted in Section 6.3, there is an interesting physical example of a spontaneously broken global SU(2) symmetry, the $\mathrm{SU(2)_{f5}}$ symmetry of $\mathcal{L}_{\mathrm{QCD}}$, in which the three massless modes are identified with the pions. We cannot consider this in any more detail here, however, being concerned rather to proceed to the local version of the SU(2)×U(1) model of (239). Such an extension is easily written down, just by using the SU(2) covariant form (163) and the U(1) covariant derivative of the form (163). In the notation we shall use in the next section, this means replacing (239) by

$$\mathcal{L}_{\mathrm{G\Phi}}=(D_\mu\phi)^\dagger(D^\mu\phi)+\mu^2\phi^\dagger\phi-\frac{\lambda}{4}(\phi^\dagger\phi)^2-\frac{1}{4}\boldsymbol{F}_{\mu\nu}\cdot\boldsymbol{F}^{\mu\nu}-\frac{1}{4}G_{\mu\nu}G^{\mu\nu} \tag{252}$$

where

$$D_\mu\phi=(\partial_\mu+ig\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu/2+ig'B_\mu/2)\phi\,, \tag{253}$$

$\boldsymbol{F}_{\mu\nu}$ is as in (177), and $G_{\mu\nu}=\partial_\mu B_\nu-\partial_\nu B_\mu$. Thus the $\boldsymbol{W}$'s are the SU(2) gauge fields, and the B is the U(1) gauge field. Equation (252) is, in fact, the gauge and Higgs field sector of the Standard Model. As in the local U(1) case, the particle spectrum is most easily found by exploiting the local gauge freedom to choose the $\boldsymbol{\theta}$ fields in (250) to vanish, as in the ansatz (229): that is, we set

$$\phi=\begin{pmatrix}0\\ (v+H(x))/\sqrt{2}\end{pmatrix}. \tag{254}$$

Substituting (254) into (252) and retaining only terms which are of second order in the fields (i.e. kinetic energies or mass terms) we find

$$\begin{aligned}\mathcal{L}_{G\Phi} \;=\;& \frac{1}{2}\partial_\mu H\partial^\mu H-\mu^2H^2\\ &-\frac{1}{4}F_{1\mu\nu}F_1^{\mu\nu}+\frac{1}{8}g^2v^2W_{1\mu}W_1^\mu\\ &-\frac{1}{4}F_{2\mu\nu}F_2^{\mu\nu}+\frac{1}{8}g^2v^2W_{2\mu}W_2^\mu\\ &-\frac{1}{4}F_{3\mu\nu}F_3^{\mu\nu}-\frac{1}{4}G_{\mu\nu}G^{\mu\nu}+\frac{1}{8}v^2(gW_{3\mu}-g'B_\mu)(gW_3^\mu-g'B^\mu)\,.\end{aligned} \tag{255}$$

The first line of (255) tells us that we have a scalar field of mass $\sqrt{2}\mu$ (the Higgs boson, again). The next two lines tell us that the components W_1 and W_2 of the triplet (W_1,W_2,W_3) acquire a mass

$$M_1=M_2=gv/2\equiv M_W\,. \tag{256}$$

The last line shows us that the fields W_3 and B are mixed. But they can easily be unmixed by noting that the last term in (255) involves only the combination $gW_3 - g'B$, which evidently acquires a mass. This suggests introducing the linear combinations

$$Z^\mu = \cos\theta_W W_3^\mu - \sin\theta_W B^\mu \tag{257}$$

$$A^\mu = \sin\theta_W W_3^\mu + \cos\theta_W B^\mu \tag{258}$$

where

$$\cos\theta_W = g/(g^2 + g'^2)^{1/2}, \quad \sin\theta_W = g'/(g^2 + g'^2)^{1/2}\,. \tag{259}$$

We then find that the last line of (255) becomes

$$-\frac{1}{4}F_{Z\mu\nu}F_Z^{\mu\nu} + \frac{1}{8}v^2(g^2 + g'^2)\,Z_\mu Z^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{260}$$

where

$$F_{Z\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu \quad \text{and} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,. \tag{261}$$

Thus

$$M_Z = \frac{1}{2}v(g^2 + g'^2)^{1/2} = M_W/\cos\theta_W \tag{262}$$

and

$$M_A = 0\,. \tag{263}$$

Counting degrees of freedom as in the local U(1) case, we originally had 12 in (252)—three massless W's and one massless B, which is 8 in all, together with 4 ϕ-fields. After symmetry breaking, we have 3 massive vector fields W_1, W_2 and Z making 9 degrees of freedom, one massless vector field A with 2, and one massive scalar H. Of course, the physical application will be to identify the W and Z fields with those physical particles, and the A field with the massless photon. In the gauge (254), the W and Z particles have propagators of the form (231).

The identification of A^μ with the photon field is made clearer if we look at the form of $D_\mu\phi$ written in terms of A_μ and Z_μ, discarding the W_1, W_2 pieces:

$$D_\mu\phi = \left\{\partial_\mu + ig\sin\theta_W\left(\frac{1+\tau_3}{2}\right)A_\mu + \frac{ig}{\cos\theta_W}\left[\frac{\tau_3}{2} - \sin^2\theta_W\left(\frac{1+\tau_3}{2}\right)\right]Z_\mu\right\}\phi\,. \tag{264}$$

Now the operator $(1 + \tau_3)$ acting on $\langle 0|\phi|0\rangle$ gives zero, as observed in (248), and this is why A_μ does not acquire a mass when $\langle 0|\phi|0\rangle \neq 0$ (gauge fields coupled to *unbroken* symmetries of $\langle 0|\phi|0\rangle$ do *not* become massive). Although certainly not unique, this choice of ϕ and $\langle 0|\phi|0\rangle$ (due to Weinberg (1967)) is undoubtedly very economical and natural. The zero eigenvalue of $(1 + \tau_3)$ can be interpreted as the electromagnetic charge of the vacuum, which we would not wish to be non-zero. We would then tentatively expect the identification

$$e = g\sin\theta_W \tag{265}$$

in order to get the right 'electromagnetic D_μ' in (264).

We have at last assembled all the conceptual ingredients we need for the electroweak theory, to which we now turn.

Problems for Lecture 5

P5.1 Verify that the inverse of the bracket $[\ldots]$ in (193) is as given in (194).

P5.2 Let

$$\mathcal{L}_\phi^{(1)} = \partial_\mu\phi^\dagger\partial^\mu\phi + \mu^2\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2.$$

Set

$$\phi(x) = \frac{1}{\sqrt{2}}(v + h(x))e^{-i\theta(x)/v} \quad .$$

Show that

$$\mathcal{L}_\phi^{(1)} = \frac{1}{2}\partial_\mu h \partial^\mu h - \mu^2 h^2 + \frac{1}{2}\partial_\mu\theta\partial^\mu\theta + \frac{\mu^4}{\lambda} + \text{ non-quadratic terms}$$

(μ^4/λ is an irrelevant constant).

P5.3 Let

$$\mathcal{L}_h = [(\partial_\mu + ieA_\mu)\phi]^\dagger[(\partial^\mu + ieA^\mu)\phi] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \mu^2\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2.$$

Set $\phi = 1/\sqrt{2}(v + h(x))$. Show that

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h\partial^\mu h - \mu^2 h^2 + \frac{\mu^4}{\lambda} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu + \text{non-quadratic terms.}$$

So $m_A = ev$.

P5.4 Let

$$\mathcal{L}_\phi^{(2)} = (\partial\phi)^\dagger(\partial^\mu\phi) + \mu^2(\phi^\dagger\phi) - \frac{\lambda}{4}(\phi^\dagger\phi)^2$$

where $\phi = \binom{\phi^+}{\phi^0}$ and $(\phi^+)^\dagger = \phi^-$, $(\phi^0)^\dagger = \bar{\phi}^0$.

Set

$$\phi = \exp(-i\boldsymbol{\theta}(x)\cdot\boldsymbol{\tau}/v)\begin{pmatrix}0\\ \frac{v+\sigma(x)}{\sqrt{2}}\end{pmatrix}.$$

Show that

$$\mathcal{L}_\phi^{(2)} = \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - \mu^2\sigma^2 + \frac{1}{2}\partial_\mu\boldsymbol{\theta}\cdot\partial^\mu\boldsymbol{\theta} + \text{non-quadratic terms.}$$

7 THE ELECTROWEAK THEORY

See Chapter 22 of Ref. [2].

We have seen that the original four-fermion theory of weak interactions is non-renormalizable, and useful only at energies well below 100 GeV. Replacing the four-fermion coupling by a Yukawa-like coupling to massive W's and Z's gave us a theory with a dimensionless coupling constant, but it was not renormalizable either. In fact, the only known way of getting a renormalizable theory of massive charged vector bosons is to regard them as gauge quanta of a spontaneously broken gauge theory. This necessitates the existence of a scalar field, the Higgs field, three of whose components correspond to the longitudinal components of the $W^\pm$ and Z^0, and the fourth of which survives as a scalar particle in the physical spectrum, but of unknown mass. In a sense, the mass of the Higgs boson m_H acts like a cut-off; but we shall see that there are quite persuasive reasons to think that at least the simplest Higgs sector model of Section 6.5 does not make sense for m_H much beyond 500–1000 GeV.

7.1 The electroweak theory for one fermion family

So far, in Section 6.5, we have only introduced the gauge and Higgs field sectors of the electroweak theory; we now need to include the quarks and leptons. Here the crucial new phenomenological input is that the weak interactions violate parity (while the electromagnetic ones of course do not). This means that the weak interaction is different for the left-handed components of fermion fields and for right-handed components. Electroweak interactions are described by a gauge theory based on a spontaneously broken local $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ invariance. The 'L' means that the SU(2) part (with the gauge fields $\boldsymbol{W}^\mu$ of Section 6.5) acts only on the left-handed parts ψ_L of fermion fields (see problem P4.1); it is therefore 'maximally' parity violating. The U(1) part (with the gauge field B^μ) acts on both right-(if any)

and left-handed components, in such a way that the particular combination (258) conserves parity, as is required for the electromagnetic interaction; the other combination (257), which mediates neutral weak interactions, will turn out not to couple in the 'pure V−A' form, as is indeed observed. The simplest structure allowing connection between the parity violating weak force and the parity conserving electromagnetic one is the $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ one, originally proposed by Glashow [9], with brave disregard for the non-renormalizability problem. The $\mathrm{SU(2)_L}$ part is often called 'weak isospin' and the U(1) 'weak hypercharge'.

In this theory, the basic fields are fermions (leptons and quarks), gauge bosons, and Higgs fields. The left-handed parts of the fermion fields form (weak isospin) doublets under $\mathrm{SU(2)_L}$

$$\text{`}\psi_{\mathrm{L}}\text{'} = \begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}u\\ \tilde{d}\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}c\\ \tilde{s}\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}t\\ tildeb\end{pmatrix}_{\mathrm{L}}, \tag{266}$$

where the $\tilde{}$ denotes states which are mixed with respect to the strong interaction states d, s and b (see the following section, and note that the colour labels will be suppressed throughout), while the right-handed components are $\mathrm{SU(2)_L}$ singlets

$$\text{`}\psi_{\mathrm{R}}\text{'} = e^-_{\mathrm{R}},\ \ \mu^-_{\mathrm{R}}, \ldots, \tag{267}$$

where for simplicity we shall generally assume in this section that the neutrinos are massless (see also Section 7.2). We shall confine the discussion in the present section to just one 'family', comprising ν_e, e^-, u and d (which should really be $\tilde{d}$ but we are ignoring mixing for the moment).

The Lagrangian can be looked at in many ways, but we shall write it as

$$\mathcal{L} = \mathcal{L}_{\mathrm{S}} + \mathcal{L}_{\mathrm{SB}} \tag{268}$$

where S stands for 'symmetrical' under $\mathrm{SU(2)}\times\mathrm{U(1)}$ and SB stands for 'symmetry breaking'. In $\mathcal{L}_{\mathrm{S}}$ we have a gauge invariant Lagrangian $\mathcal{L}_{\mathrm{f}}$ describing the interactions of the fermions with the $\boldsymbol{W}$ and B fields, together with the SU(2) Yang–Mills Lagrangian $\mathcal{L}_W$ (179) for the $\boldsymbol{W}$ fields and the U(1) Lagrangian $\mathcal{L}_B$ for the B field as in (252); in $\mathcal{L}_{\mathrm{SB}}$ we shall have everything involving the Higgs fields. In Section 4.2 we learned how to construct a locally SU(2) invariant gauge theory with a fermion doublet (see (163)). The difference now is that we want the $\mathrm{SU(2)_L}$ to act only on the L-component of the doublet. However, there is no problem with this for *massless* fields: (218) shows us that the 'kinetic' operator $\not{\partial}$ does not mix L and R components, and hence there is no objection to 'gauging' each of them differently (i.e. using a different $\not{D}$ on ψ_{L} and on ψ_{R}). On the other hand, (219) shows that this is *not* true for the mass terms—a difficulty we shall deal with shortly by getting the mass terms from $\mathcal{L}_{\mathrm{SB}}$. First, we simply state that the appropriate D's are in fact

$$D_\mu \;=\; \partial_\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu/2 + ig'yB_\mu/2 \quad \text{on } \psi_{\mathrm{L}}\text{'s} \tag{269}$$

and

$$D_\mu \;=\; \partial_\mu + ig'yB_\mu/2 \quad \text{on } \psi_{\mathrm{R}}\text{'s}\,, \tag{270}$$

where the condition

$$Q = \tau_3/2 + y/2 \tag{271}$$

is imposed, Q being the electric charge in units of e (the positron charge). The factor of $\frac{1}{2}$ in the B-term of (269) is conventional, but (271) fixes the normalization of the coupling g'. The eigenvalues of the $\tau_3/2$ operator in (269) are as indicated by the placings in (266): namely $+\frac{1}{2}$ for $(\nu_e, \nu_\mu, \nu_\tau, u, c, s)_{\mathrm{L}}$ and $-\frac{1}{2}$ for e^-_{L}, etc. For the $(\text{lepton})_{\mathrm{L}}$ fields the y eigenvalue is -1, while for the $(\text{quark})_{\mathrm{L}}$ fields it is $+\frac{1}{3}$; for the R-fields y is just $2Q$ since the $\tau_3/2$ eigenvalue is zero.

The gauge invariant Lagrangian $\mathcal{L}_{\mathrm{f}}$ (for massless fermions) is therefore

$$\mathcal{L}_{\mathrm{f}} \;=\; \overline{\ell}_{e\mathrm{L}} i\not{D}\ell_{e\mathrm{L}} + \overline{q}_{\mathrm{L}} i\not{D} q_{\mathrm{L}} + \overline{e}_{\mathrm{R}} i\not{D} e_{\mathrm{R}} + \overline{u}_{\mathrm{R}} i\not{D} u_{\mathrm{R}} + \overline{d}_{\mathrm{R}} i\not{D} d_{\mathrm{R}} \tag{272}$$

where

$$\ell_{e\mathrm{L}} = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_{\mathrm{L}}, \quad q_{\mathrm{L}} = \begin{pmatrix} u \\ d \end{pmatrix}_{\mathrm{L}} \tag{273}$$

and a $\nu_{e\mathrm{R}}$ term can be added to (272) if desired. From (272) we can already read off the couplings of the charged W's to the fermions (the W_3 and B will mix, as we saw in Section 6.5). The correct normalization for charged fields is that $W^\mu = (W_1 - iW_2)/\sqrt{2}$ destroys the W^+ or creates W^-, so that the $\boldsymbol{\tau}\cdot\boldsymbol{W}/2$ terms are

$$\frac{1}{\sqrt{2}}\left\{\tau_+ \frac{(W_1 - iW_2)}{\sqrt{2}} + \tau_- \frac{(W_1 + iW_2)}{\sqrt{2}}\right\} + \tau_3 \frac{W_3}{2} \tag{274}$$

where $\tau_\pm = (\tau_1 \pm i\tau_2)/2$ are the raising and lowering operators for the doublet. Thus the first term in (274) picks out the process $e^- \to \nu_e W^-$ for example, with the result that the corresponding vertex is

$$-\frac{ig}{\sqrt{2}}\gamma_\mu \frac{(1-\gamma_5)}{2}, \tag{275}$$

and similarly for the quarks (if unmixed), and other families. Hence we can immediately make a connection with the original V–A Fermi theory of these charged current processes, namely

$$G_{\mathrm{F}}/\sqrt{2} = g^2/8M_W^2 . \tag{276}$$

Although the quark couplings can also be read off from (272), they are unphysical at this stage since mixing has not yet been introduced.

There are also couplings of the Z^0 to fermions. To find these, we need to rewrite the neutral part of the D's in (269) and (270) in terms of the Z and A fields defined in (257) and (258) [cf. (264)]. We find

$$D_\mu(\mathrm{neutral}) = \partial_\mu + ieQA_\mu + \frac{igZ_\mu}{2\cos\theta_W}(v_f - a_f\gamma_5) \tag{277}$$

where

$$v_f = \frac{\tau_3}{2} - 2Q\sin^2\theta_W \tag{278}$$

and

$$a_f = \frac{\tau_3}{2} . \tag{279}$$

We see that, as remarked earlier, the Z (or 'neutral-current') coupling is not pure V–A. The Z-couplings analogous to (275) are therefore

$$\frac{-ig}{2\cos\theta_W}\gamma_\mu(v_f - a_f\gamma_5) . \tag{280}$$

The coupling (280) is that observed around the Z^0 peak.

We may write effective four-fermion interactions (valid for energies much less than M_W, M_Z) as

$$\frac{G_{\mathrm{F}}}{\sqrt{2}}\, j^C_{\mu+}\, j^{C\mu}_- \tag{281}$$

for the charged current processes, with

$$j^C_{\mu\pm} = (\overline{\psi}_2\gamma_\mu(1-\gamma_5)\tau_\pm\psi_1) , \tag{282}$$

and as

$$\sqrt{2}G_{\mathrm{F}}\rho\, j^N_\mu\, j^{N\mu} \tag{283}$$

for the neutral current processes, where

$$j^N_\mu = \overline{\psi}_f\gamma_\mu(v_f - a_f\gamma_5)\psi_f \tag{284}$$

and the quantity

$$\rho = M_W^2/M_Z^2 \cos^2\theta_W \tag{285}$$

has the value 1 in the Standard Model, at tree level.

The vector boson masses arise through symmetry breakdown via the Higgs sector, in the Standard Model, as discussed in Section 6.5. After spontaneous symmetry breaking, we have

$$\begin{aligned} M_W &= gv/2 = \cos\theta_W M_Z && (286)\\ \cos\theta_W &= g/(g^2+g'^2)^{1/2} && (287)\\ e &= g\sin\theta_W && (288)\\ m_H &= \sqrt{2}\,\mu && (289) \end{aligned}$$

in terms of the fundamental coupling parameters g, g' of the SU(2)×U(1) gauge group, and the parameters v and μ of the Higgs potential. There is also the low-energy connection (276), which we can write as

$$\frac{v}{\sqrt{2}} = 2^{-3/4} G_{\rm F}^{-1/2} = 174.1\ {\rm GeV}\,, \tag{290}$$

using $G_{\rm F} \approx 1.17 \times 10^{-5}\ {\rm GeV}^{-2}$. This gives us the scale of $\langle 0|\phi|0\rangle$, for which as yet there is no theoretical explanation. We may also write (276) as

$$\begin{aligned} M_W &= (\pi\alpha/\sqrt{2}\,G_{\rm F})^{1/2}/\sin\theta_W && (291)\\ &= 37.2802\ {\rm GeV}/\sin\theta_W && (292) \end{aligned}$$

using the conventional low-energy value of α. Note that all the above relations are between parameters in the Lagrangian, and hold at the tree level only; they can be changed by loop corrections (see Section 7.4).

We must now consider how to bring fermion masses into this theory. We begin by noting, again, that a typical Dirac mass term has the form (219), which is clearly not invariant under transformations which treat $\psi_{\rm L}$ and $\psi_{\rm R}$ differently. Would it matter if we just added in such a mass term? The answer is that if we did this the theory would, once again, not be renormalizable. And, once again, we can arrange for the fermions to 'acquire mass spontaneously', this time via couplings of the generic 'Yukawa' type $g_f\bar\psi\psi\phi$. This can be made SU(2)$_{\rm L}$×U(1) invariant, and then if the scalar field acquires a vacuum value v we have a mass term (in such a vacuum) equal to $g_f v$. Some such treatment of fermion masses is necessary for the theory to make sense much beyond the W–Z mass range.

It is obviously most economical if we can 'blame' fermion masses on the same Higgs field that generates the W and Z masses, but it must be recognized that the Yukawa coupling 'mechanism' is on a very different footing from the symmetry-inspired gauge couplings—at least in the absence of any further symmetry that might relate these two types of coupling. At any rate, consider the case of the ν_e, e^- doublet, in the simple case that the ν_e is massless, with a Yukawa coupling between these fields and the standard doublet Higgs, of the type

$$-g_e(\overline{\ell}_{e\rm L}\phi e_{\rm R} + \overline{e}_{\rm R}\phi^\dagger \ell_{e\rm L})\,. \tag{293}$$

Remembering that $e_{\rm R}$ is an SU(2) scalar, we see that (293) is Lorentz invariant, and invariant under global SU(2) transformations (because $\bar\ell\phi$ and $\phi^\dagger\ell$ are invariant); it is also invariant under U(1)$_y$ transformations, with the y assignments made after (271), if $y(\phi) = 1$ (which is what we actually assumed in (253)). In fact, since no derivatives are involved in (293), it is also invariant under local SU(2)×U(1) transformations. But the Higgs sector contains the potential $V(\phi)$ of (239), which 'triggers' spontaneous symmetry breaking. The vacuum value (246) for ϕ when inserted into (293), yields

$$-(g_e v/\sqrt{2})(\overline{e}_{\rm L}e_{\rm R} + \overline{e}_{\rm R}e_{\rm L}) \tag{294}$$

which is precisely a mass term for the electron if we identify

$$g_e = m_e\sqrt{2}/v\,. \tag{295}$$

When oscillations about this vacuum are considered, in the simple gauge of (254), one easily finds that the H-field couples to the electron with a vertex

$$-im_e/v\,. \tag{296}$$

Sure enough, the coupling is proportional to the electron mass—and on dimensional grounds to v^{-1}.

It might seem from the foregoing that only a mass for the $t_3 = -\frac{1}{2}$ component of the fermion doublets could be generated this way, because of the form of $\langle 0|\phi|0\rangle$. Remarkably enough, however, the same Higgs field can also provide a mass for the $t_3 = +\frac{1}{2}$ component (and this is of course necessary for the quarks, if not for the neutrinos). It can be shown that the field ϕ_c defined by

$$\phi_c = i\tau_2\phi^* = \begin{pmatrix} (\phi_3 + i\phi_4)/\sqrt{2} \\ -(\phi_1 + i\phi_2)/\sqrt{2} \end{pmatrix} = \begin{pmatrix} \overline{\phi}^0 \\ -\phi^- \end{pmatrix}, \tag{297}$$

where (243) has been used, is also an isodoublet. (The notation in (297) is reminiscent of the K-meson doublet $(\bar{K}^0, K^-)$; alternatively, we may think of a quark isospin doublet like $\begin{pmatrix} u \\ d \end{pmatrix}$ and its conjugate doublet $\begin{pmatrix} \bar{d} \\ -\bar{u} \end{pmatrix}$, with the $I = 0$ combination being $(\bar{d}d - \bar{u}u)$.) With the help of ϕ_c we can write down another gauge invariant coupling in the ν_e-e sector, namely

$$-g_{\nu_e}\left(\overline{\ell}_{e\mathrm{L}}\phi_c\nu_{e\mathrm{R}} + \overline{\nu}_{e\mathrm{R}}\phi_c^\dagger\ell_{e\mathrm{L}}\right) \tag{298}$$

which produces

$$-\left(g_{\nu_e}v/\sqrt{2}\right)\left(\overline{\nu}_{e\mathrm{L}}\nu_{e\mathrm{R}} + \overline{\nu}_{e\mathrm{R}}\nu_{e\mathrm{L}}\right) \tag{299}$$

in the Higgs vacuum (246), which is a neutrino mass term (if required) provided $g_{\nu_e} = \sqrt{2}m_{\nu_e}/v$. Once again, the H-field will couple with an amplitude of the form (296), with $m_e \to m_{\nu_e}$. The procedure can obviously be repeated for the u and d quarks.

It is clearly possible to go on like this, and arrange for as many fermion families to have a mass as is required—and we will look at this a little more closely in the next section. However, one must note that the theory does no more than accommodate itself to the mass difficulty: in no sense do the fermion masses 'come out' of the theory, since each has simply to be inserted by hand via a new Yukawa coupling. In essence, these Yukawa couplings are *not* gauge interactions, and hence not universal.

The Higgs coupling to fermions can now be written generally as

$$-iem_f/2\sin\theta_W M_W\,. \tag{300}$$

There are also trilinear and quadrilinear Higgs self-couplings arising from the $\lambda(\phi^\dagger\phi)^2$ term in (252). Recalling that $\lambda = 4\mu^2/v^2$ and that $m_H = \sqrt{2}\mu$, we can write the trilinear coupling as

$$-i3m_H^2 e/8M_W\sin\theta_W \tag{301}$$

and the quadrilinear as

$$-i3m_H^2e^2/16M_W^2\sin^2\theta_W\,. \tag{302}$$

There are also the trilinear H–W^+–W^-

$$ieM_W g_{\lambda\mu}/\sin\theta_W \tag{303}$$

and H–Z–Z

$$i2eM_Z g_{\lambda\mu}/\sin 2\theta_W \tag{304}$$

couplings, together with quadrilinear ϕ^2W^2, ϕ^2Z^2 couplings which we shall not give here. Note that all these couplings are determined by the existing set of parameters—and, in particular, that the Higgs couples most strongly to the heaviest particles, so that decays to heavy channels offer the largest rates.

7.2 The three-family model

We now extend the preceding discussion to the three-family case, which will involve the important subjects of quark flavour mixing in charged current processes (and of no mixing—the GIM mechanism (Glashow et al. [10])—in neutral current processes), and CP violation. We shall here assume that there are just three families. We introduce three doublets of left-handed fields

$$q_{\mathrm{L}1} = \begin{pmatrix} u_{\mathrm{L}1} \\ d_{\mathrm{L}1} \end{pmatrix}, \quad q_{\mathrm{L}2} = \begin{pmatrix} u_{\mathrm{L}2} \\ d_{\mathrm{L}2} \end{pmatrix}, \quad q_{\mathrm{L}3} = \begin{pmatrix} u_{\mathrm{L}3} \\ d_{\mathrm{L}3} \end{pmatrix} \tag{305}$$

and the corresponding six singlets

$$u_{\mathrm{R}1}, \quad d_{\mathrm{R}1}, \quad u_{\mathrm{R}2}, \quad d_{\mathrm{R}2}, \quad u_{\mathrm{R}3}, \quad d_{\mathrm{R}3}\,, \tag{306}$$

which transform in the now familiar way under $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)$. The u-fields correspond to the $t_3 = +\frac{1}{2}$ components of $\mathrm{SU}(2)_{\mathrm{L}}$, the d ones to the $t_3 = -\frac{1}{2}$ components, and to their 'R' partners. The labels 1, 2, and 3 refer to the family number; for example, with no mixing at all, $u_{\mathrm{L}1} = u_{\mathrm{L}}$, $d_{\mathrm{L}1} = d_{\mathrm{L}}$, etc. (We are thinking of (305) and (306) as quark fields, but the discussion will be quite general and could just as well apply to leptons if they should need mixing too—we return to leptons later.) We have to consider what is the most general $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)$-invariant interaction between the Higgs field (assuming we can still get by with only one) and these various fields. Apart from the symmetry, the only other theoretical requirement is renormalizability—for, after all, if we drop this we might as well abandon the whole motivation for the 'gauge' concept. This implies (as in the discussion of the Higgs potential V) that we cannot have terms like $(\bar{\psi}\psi\phi)^2$ appearing—which would have a coupling with dimensions $(\text{mass})^{-4}$ and would be non-renormalizable. In fact the only renormalizable Yukawa coupling is of the form '$\bar{\psi}\psi\phi$', which has a dimensionless coupling (as in the g_e and g_{ν_e} of (293) and (298)). However, there is no *a priori* requirement for it to be 'diagonal' in the weak interaction family index i. The allowed generalization of (293) and (298) is therefore an interaction of the form (summing on repeated indices)

$$\mathcal{L}_{\psi\phi} = a_{ij}\bar{q}_{\mathrm{L}i}\phi^c u_{\mathrm{R}j} + b_{ij}\bar{q}_{\mathrm{L}i}\phi d_{\mathrm{R}j} + \mathrm{h.c.} \tag{307}$$

where

$$q_{\mathrm{L}i} = \begin{pmatrix} u_{\mathrm{L}i} \\ d_{\mathrm{L}i} \end{pmatrix} \tag{308}$$

and a sum on the family indices i and j (from 1 to 3) in (307) is assumed. After symmetry breaking, using the gauge (254), we find

$$\mathcal{L}_{f\phi} = -\left(1 + \frac{H}{v}\right)\left[\bar{u}_{\mathrm{L}i}m^u_{ij}u_{\mathrm{R}j} + \bar{d}_{\mathrm{L}i}m^d_{ij}d_{\mathrm{R}j} + \mathrm{h.c.}\right] \tag{309}$$

where the 'mass matrices' are

$$m^u_{ij} = -\frac{v}{\sqrt{2}}\,a_{ij}, \quad m^d_{ij} = -\frac{v}{\sqrt{2}}\,b_{ij}\,. \tag{310}$$

Although we have not indicated it, the m^u and m^d matrices could involve a 'γ_5' part as well as a '1' part in Dirac space. It can be shown (Weinberg [11], Feinberg et al. [12]) that m^u and m^d can both be made Hermitean, γ_5-free, and diagonal by making four separate unitary transformations on the 'family triplets'

$$u_{\mathrm{L}} = \begin{pmatrix} u_{\mathrm{L}1} \\ u_{\mathrm{L}2} \\ u_{\mathrm{L}3} \end{pmatrix}, \quad d_{\mathrm{L}} = \begin{pmatrix} d_{\mathrm{L}1} \\ d_{\mathrm{L}2} \\ d_{\mathrm{L}3} \end{pmatrix}, \ \text{etc.} \tag{311}$$

via

$$u_{\mathrm{L}\alpha} = \left(U^{(u)}_{\mathrm{L}}\right)_{\alpha i} u_{\mathrm{L}i}, \quad u_{\mathrm{R}\alpha} = \left(U^{(u)}_{\mathrm{R}}\right)_{\alpha i} u_{\mathrm{R}i}\ , \tag{312}$$

$$d_{\mathrm{L}\alpha} = \left(U_{\mathrm{L}}^{(d)}\right)_{\alpha i} d_{\mathrm{L}i}, \quad d_{\mathrm{R}\alpha} = \left(U_{\mathrm{R}}^{(d)}\right)_{\alpha i} d_{\mathrm{R}i} \ . \tag{313}$$

In this notation, 'α' is the index of the 'mass diagonal' basis, and 'i' is the 'weak interaction' basis. Then (309) becomes

$$\mathcal{L}_{q\psi} = -\left(1+\frac{H}{v}\right)\left[m_u\bar{u}u + \cdots + m_b\bar{b}b\right] \ . \tag{314}$$

Rather remarkably, we can still manage with only the one Higgs field. It couples to each fermion with a strength proportional to the mass of that fermion, divided by M_W.

Now consider the $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ gauge invariant interaction part of the Lagrangian. Written out in terms of the 'weak interaction' fields $u_{\mathrm{L,R}\,i}$ and $d_{\mathrm{L,R}\,i}$ [cf. (269) and (270)], it is

$$\begin{aligned}\mathcal{L}_{fW,B} &= i\left(\bar{u}_{\mathrm{L}j}, \bar{d}_{\mathrm{L}j}\right)\gamma^\mu(\partial_\mu + ig\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu/2 + ig'yB_\mu/2)\begin{pmatrix}u_{\mathrm{L}j}\\ d_{\mathrm{L}}j\end{pmatrix} \\ &\quad + i\bar{u}_{\mathrm{R}j}\gamma^\mu(\partial_\mu + ig'yB_\mu/2)u_{\mathrm{R}j} + i\bar{d}_{\mathrm{R}j}\gamma^\mu(\partial_\mu + ig'yB_\mu/2)d_{\mathrm{R}j}\end{aligned} \tag{315}$$

where a sum on j is understood. This now has to be rewritten in terms of the mass-eigenstates $u_{\mathrm{L,R}\,\alpha}$ and $d_{\mathrm{L,R}\,\alpha}$.

Problem P6.1 shows that the neutral current part of (315) is diagonal in the mass basis—that is, the neutral current interactions do not change the flavour of the physical (mass eigenstates) quarks. The charged current processes, however, involve the *non*-diagonal matrices τ_1 and τ_2 in (315), and this spoils the argument used in problem P6.1. Indeed, using (274) we find that the charged current piece is

$$\begin{aligned}\mathcal{L}_{\mathrm{cc}} &= -\frac{g}{\sqrt{2}}(\bar{u}_{\mathrm{L}j}, d_{\mathrm{L}j})\gamma_\mu\tau_+ W_\mu\begin{pmatrix}u_{\mathrm{L}j}\\ d_{\mathrm{L}j}\end{pmatrix} + \mathrm{h.c.} \\ &= -\frac{g}{\sqrt{2}}\bar{u}_{\mathrm{L}j}\gamma^\mu d_{\mathrm{L}j}W_\mu + \mathrm{h.c.} \\ &= -\frac{g}{\sqrt{2}}\bar{u}_{\mathrm{L}\alpha}\left[\left(U_{\mathrm{L}}^{(u)\,\dagger}\right)_{\alpha i}\left(U_{\mathrm{L}}^{(d)}\right)_{i\beta}\right]\gamma^\mu d_{\mathrm{L}\beta}W_\mu + \mathrm{h.c.}\end{aligned} \tag{316}$$

where the matrix

$$V_{\alpha\beta} \equiv \left[U_{\mathrm{L}}^{(u)\,\dagger}U_{\mathrm{L}}^{(d)}\right]_{\alpha\beta} \tag{317}$$

is not diagonal, though it is unitary. V therefore has 9 real parameters, which can be reduced to 4—three 'rotational angles' and one phase—by redefinitions of the quark fields (Jarlskog [13]). This is the famous CKM matrix, (Cabibbo [14], Kobayashi and Maskawa [15]) the interaction (316) having the form

$$-\frac{g}{\sqrt{2}}W_\mu\left(\bar{u}_{\mathrm{L}}\ \bar{c}_{\mathrm{L}}\ \bar{t}_{\mathrm{L}}\right)\begin{pmatrix}V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix}\begin{pmatrix}d_{\mathrm{L}}\\ s_{\mathrm{L}}\\ b_{\mathrm{L}}\end{pmatrix} + \mathrm{h.c.} \tag{318}$$

The entries in the V-matrix modify the vertex (275) in an obvious way. The single phase δ in the V-matrix accommodates CP-violation. In the case of only two flavours, V has only 1 real parameter, which is the Cabibbo angle, and there is no freedom to have a CP violation phase in the family mixing matrix. It is an important challenge to experiment to find out whether all CP-violating phenomena can be described with just this one parameter δ in the CKM matrix (see the lectures on CP violation).

Returning finally to the leptons, all of the above will apply (with three more mixing angles and one more phase) if the neutrinos do in fact have a mass. We would then have leptonic flavour mixing in c.c. processes, involving a term of the form $[\bar{\ell}_{\mathrm{L}}V_\ell\gamma^\mu\ell'_{\mathrm{L}}W_\mu + \mathrm{h.c.}]$ [cf. (316)], and lepton mass terms $[\bar{\ell}_{\mathrm{L}}m_\ell\ell_{\mathrm{R}} + \mathrm{h.c.}]$ and $[\bar{\ell}'_{\mathrm{L}}m'_\ell\ell'_{\mathrm{R}} + \mathrm{h.c.}]$, where V_ℓ is the leptonic analogue of (317), and m_ℓ, m'_ℓ are the analogues of the quark masses. There is nothing in the Standard Model that requires the neutrinos to be massless, and indeed the experimental data now imply that more than one is not; in GUTs they generally do have (small) masses—see the lectures on neutrinos.

7.3 One remark about the Higgs sector

The Higgs sector is the one big unknown still hanging over the Standard Model, starting with the question: What is the Higgs mass? There is an interesting theoretical argument here which is worth a mention.

We first note that, for a given vacuum value v as in (290), the Higgs mass is [cf. (241) and (251)]

$$m_H = v\lambda^{\frac{1}{2}}/\sqrt{2} \sim \lambda^{\frac{1}{2}} \times 174 \text{ GeV} \ . \tag{319}$$

Now λ is a dimensionless constant: if it is $O(\alpha)$ we would say that the theory is perturbative, while if it is $O(1)$ we would say it was strongly coupled. It is clear from (319), and the present experimental lower bounds on m_H, that we are already not far from the strongly coupled region. But we can ask: Can λ (the renormalized coupling) take *any* value at all? That is, can m_H (for fixed v) be arbitrarily large?

To answer this we must recall that, in a renormalizable theory, 'the' value of λ has to be defined at a certain scale, and the value at another scale is different (i.e. λ 'runs'). For the interaction (239), calculation shows that the analogue of (129) is

$$\lambda(E) = \lambda \Big/ \left[1 - \frac{3}{8\pi^2}\lambda \ln\left(\frac{E}{v}\right)\right] \tag{320}$$

taking the 'physical' λ to be defined at the scale v. Note that this theory, like QED, is *not* asymptotically free. It follows from (320) that the theory breaks down (or, more conservatively, $\lambda(E)$ becomes so large that all perturbative expectations are useless) at an energy E^* such that $E^* \sim v \exp(\frac{8\pi^2}{3\lambda})$. But, for given v, we also have from (319) that λ is related to m_H. So the theory breaks down at

$$E^* \sim v \exp(\frac{4\pi^2 v^2}{3m_H^2}) \ . \tag{321}$$

This is a very remarkable formula, because it is exponentially sensitive to the unknown m_H—and it is particularly interesting that the Higgs mass is in the denominator of the exponent. For 'small' m_H the breakdown scale is high—e.g. for $m_H \sim 150$ GeV, $E^* \sim 6 \times 10^{17}$ GeV. But for $m_H \sim 700$ GeV, E^* is already as low as 1 TeV. Clearly, at such a value of m_H, the Higgs mass is essentially equal to the 'breakdown scale' itself, and m_H cannot get any higher without new physics intervening in one form or another: maybe non-perturbative phenomena, or maybe supersymmetry.

7.4 Two remarks on one-loop corrections in the Standard Model

The precision of LEP and other data (of order 0.1%) was such that the measurements were sensitive to one-loop effects—and the very high quality of the fits to all the data confirm the presence of these corrections very convincingly. What is particularly interesting is that the loop corrections could be used to make *predictions* about as yet unseen particles: for example, the top quark mass was predicted to be something like 175 ± 10 GeV via its virtual effects in loops, *before* it was discovered as a real particle! (and the errors on the experimental mass determination were similar!). A typical fit to all data (Grünewald [16]) has a χ^2/d.o.f of 14.9/15, corresponding to a probability of 46%. This extremely strong numerical consistency lends impressive support to the belief that we are indeed dealing with a renormalizable spontaneously broken gauge theory, because *no extra parameters, not in the original Lagrangian, have had to be introduced*. In fact, one can turn this around. It is widely believed that, remarkably successful as it is, the Standard Model is not the end of physics, and that consequently further parameters will be required at some stage. The close agreement between the data and the existing Standard Model means that the new physics is proving very hard to see, at present energies.

As we have seen, we obtain cut-off independent results from loop corrections in a renormalizable theory by taking certain parameters (those appearing in the original Lagrangian) from experiment. In the electroweak case, it is usual to take the set

$$\alpha,\ G_F,\ m_Z,\ m_H,\ m_f, \text{parameters of mixing matrices;} \tag{322}$$

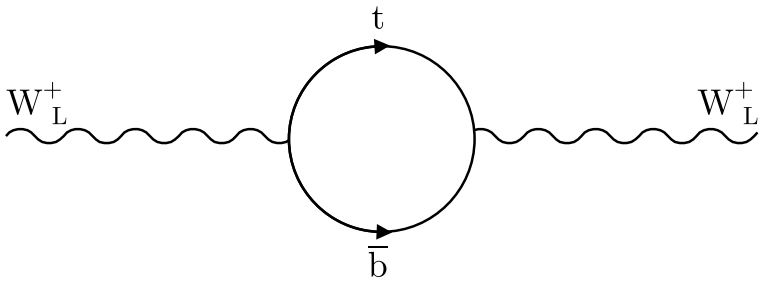


Fig. 17: *t–b* vacuum polarization loop.

(α_s of QCD and the QCD θ-parameter need to be added for the full Standard Model). After renormalization, one can derive radiatively-corrected values for physical quantities in terms of the set (322). For example, the tree-level relation (291) takes the following form at one loop:

$$M_W^2 = \left[(\pi\alpha/\sqrt{2}\,G_{\rm F})/\sin^2\theta_W\right]/(1-\Delta r) \tag{323}$$

where $\sin\theta_W$ has been *defined* as $\sin^2\theta_W \equiv 1 - M_W^2/M_Z^2$. Δr is the one-loop correction.

We cannot go into all the details of Δr, but we do want to focus on two important features of the result (which are typical of other radiatively-corrected formulae). The leading terms in Δr have the form

$$\Delta r = \Delta\alpha - \cot^2\theta_W\Delta\rho + (\Delta r)_{\rm rem}\,. \tag{324}$$

In (324), $\Delta\alpha$ is precisely the quantity $\bar{\Pi}^2_\gamma(M_Z^2)$ which entered into the running QED constant α discussed in Section 5.3 [see (127) and after (129)]. $\Delta\rho$ is given by

$$\Delta\rho = \frac{3G_{\rm F}(m_t^2 - m_b^2)}{8\pi^2\sqrt{2}}\,, \tag{325}$$

while the 'remainder' $(\Delta r)_{\rm rem}$ contains a non-negligible term proportional to $\ln(m_t/M_Z)$, and a contribution from the Higgs boson which is (for $m_H \gg M_W$)

$$(\Delta r)_{{\rm rem},H} \approx \frac{\sqrt{2}\,G_F M_W^2}{16\pi^2}\frac{11}{3}\left[\ln\left(\frac{m_H^2}{M_W^2}\right) - \frac{5}{6}\right]\,. \tag{326}$$

The running of α is no surprise, but (325) and (326) contain unexpected features.

As regards (325), it is associated with top–bottom quark loops in vacuum polarization amplitudes, of the kind discussed for $\bar{\Pi}^{[2]}_\gamma$, but in weak boson propagators. In the QED case, referring to (125) we see that the contribution of very heavy fermions (e.g. the top) in a vacuum polarization loop should be suppressed, appearing as '$O(q^2/m_t^2)$'. This seems plausible enough: after all, the mass appears in the fermion propagator and hence in the denominator of the loop integral expression. Yet in fact m_f^2 appears in the *numerator* of (325)! the usual case ($\sim q^2/m^2$) is termed 'decoupling' of heavy matter, and it is certainly what we would expect intuitively; in (325) we have 'non-decoupling'.

We can understand the appearance of the fermion masses (squared) in the numerator as follows. The shift $\Delta\rho$ is associated with vector boson vacuum polarization contributions, for example the one shown in Fig. 17. Consider in particular the contribution from the longitudinal polarization components of the W's. As we have seen, these components are nothing but three of the four Higgs components which the $W^\pm$ and Z^0 'swallowed' to become massive. But the couplings of these 'swallowed' Higgs fields to fermions are determined by just the same Higgs–fermion Yukawa couplings as we introduced to generate the fermion masses via spontaneous symmetry breaking. Hence we expect the fermion loops to

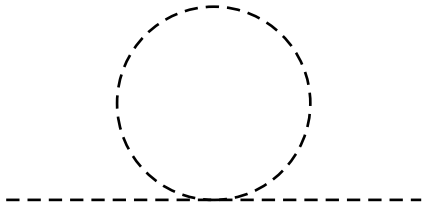


Fig. 18: One-loop self-energy graph in ϕ^4 theory.

contribute (to these longitudinal W states) something of order $g_f^2/4\pi$ where g_f is the Yukawa coupling. Since $g_f \sim m_f/v$ [see (295)] we arrive at an estimate $\sim m_f^2/4\pi v^2 \sim G_{\rm F} m_f^2/4\pi$ as in (325). An important message is that particles whose mass is proportional to their coupling to some field (i.e. in this case the Higgs field) do not 'decouple'.

But we still have to explain why $\Delta\rho$ vanishes if $m_t = m_b$. This has to do with a further symmetry of the assumed Higgs sector. As the notation suggests, $\Delta\rho$ is a leading order correction to the ρ parameter introduced in (283) and (285). At tree level, ρ has the value 1, which is a reflection of the fact that the (mass)^2 matrix, in terms of the original SU(2)$_{\rm L}\times$U(1) fields $\boldsymbol{W}^\mu$ and B^μ was [cf. (255)]

$$\frac{v^2}{4}\begin{pmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & -gg' \\ 0 & 0 & -gg' & g'^2 \end{pmatrix} \tag{327}$$

acting in the $(W_1^\mu\, W_2^\mu\, W_3^\mu\, B^\mu)$ space. Notice now that the leading 3×3 block of this matrix, acting on the $\boldsymbol{W}$'s alone, is proportional to the unit matrix. This would be the natural consequence of an unbroken SU(2) symmetry in which the $\boldsymbol{W}$'s form an SU(2) triplet. Now, with the doublet Higgs of the form (243), it is a striking fact that the Higgs potential only involves the (globally) SO(4)-symmetric combination

$$\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^3 \ . \tag{328}$$

The vacuum expectation value (246) singles out one of the four components, and breaks the SO(4) symmetry of the Higgs sector down to an SO(3), which is equivalent to the SU(2) of the $\boldsymbol{W}$'s, above. This (global) symmetry is called the 'custodial symmetry' of the (assumed) Higgs sector. It is this symmetry, in fact, that guarantees $\rho = 1$ *to all orders*.

However, examination of the behaviour of the quark mass terms under such global SU(2) transformations shows that the symmetry is explicitly broken by a difference in the masses of two quarks in the same doublet. This explains the 'm_t^2–m_b^2' dependence of the non-decoupled *t–b* loop correction. Phenomenologically this m_t^2 dependence was of great importance, because of course it meant that (paradoxically!) the heavier the top was, the more visible its effect in such loops would be. Its 'virtual' discovery was a wonderful cooperative achievement between theory and experiment.

The case is unfortunately 'reversed', in a sense, for the Higgs—and this is our second remark about loops. Without the Higgs particle, the Standard Model is non-renormalizable, and hence one might expect to see some radiative correction becoming large $O(m_H^2)$ as one tried to 'banish' the Higgs from the theory by sending $m_H \to \infty$ (m_H would be acting like a cut-off Λ). The reason is that in such a 'ϕ^4' theory, the simplest loop we meet is that shown in Fig. 18, and it is easy to see by counting powers as usual that it diverges as the square of the cut-off.

However, even without a Higgs contribution it turns out that the theory is renormalizable at the one-loop level for zero fermion masses (Veltman [17], [18]). Thus one suspects that the large m_H^2 effects will not be so dramatic after all. In fact, calculation shows (Veltman [19]; Chanowitz et al. [20], [21]) that one-loop radiative corrections grow at most like $\ln m_H^2$ for large m_H. While there are finite corrections

which are approximately $O(m_H^2)$ for $m_H^2 \ll M_{W,Z}^2$, for $m_H^2 \gg M_{W,Z}^2$ the $O(m_H^2)$ pieces cancel out from all observable quantities, leaving only $\ln m_H^2$ terms. This is just what we have in (326), and it means, unfortunately, that the sensitivity of the data to the last remaining parameter of the Standard Model (not counting the neutrino parameters!) is only logarithmic. Fits to data typically give m_H in the region of 100 GeV at the minimum of the χ^2 curve, but the error (which is not simple to interpret) is of the order of 50 GeV. Direct searches now rule out a Higgs mass less than about 110 GeV, while the $\sim$ 2.5 s.d. effect seen just before LEP closed down gave $m_H \sim 114$ GeV.

At the two-loop level, the expected $O(m_H^4)$ behaviour becomes $O(m_H^2)$ instead (van der Bij and Veltman [22], van der Bij [23])—and of course appears (relative to the one-loop contributions) with an additional factor of $O(\alpha)$. This relative insensitivity of the radiative corrections to m_H, in the limit of large m_H, was discovered by Veltman [19] and called a 'screening' phenomenon by him: for large m_H (which also means, as we have seen, large λ) we have an effectively strongly interacting theory whose principal effects are screened off from observables at lower energy. It was shown by Einhorn and Wudka [24] that this screening is also a consequence of the (approximate) isospin-SU(2) symmetry we have just discussed in connection with (325). Phenomenologically, the upshot is that it is unfortunately very difficult to get a good handle on the value of m_H from fits to the precision data.

Problems for Lecture 6

P6.1 Show that the neutral current couplings are diagonal in the 'mass' basis.

P6.2 Suppose that we took the Higgs field to be a triplet of $\mathrm{SU(2)_L}$ instead of a doublet; and suppose $\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}$ in the gauge in which it is real. The non-vanishing component has $t_3 = -1$, using

$$t_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

in the familiar 'spherical' basis. Since we want the charge of the vacuum to be zero ($Q|0\rangle = 0$) and $Q = t_3 + y/2$, we need to pick $y(\phi) = 2$. So the covariant derivative on ϕ is

$$(\partial_\mu + ig\boldsymbol{t}\cdot\boldsymbol{W}^\mu - ig'B^\mu)\phi$$

where

$$t_1 = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad t_2 = \begin{pmatrix} 0 & \frac{-i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{-i}{\sqrt{2}} \\ 0 & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}$$

and t_3 is as above (this is the more familiar set of three matrices satisfying $[t_1, t_2] = it_3$, a change of basis from the set $(t_i)_{jk} = -i\epsilon_{ijk}$). Show that the photon and Z fields are still (257) and (258), with the same $\sin\theta_W$ as in (259), but that now

$$M_Z = \sqrt{2}M_W/\cos\theta_W.$$

What would be the parameter ρ, at tree level, for this model?

ACKNOWLEDGEMENTS

It is a very great pleasure to thank all those responzible for organizing this School, and for making it so successful and memorable. In particular, I should like to thank George Pogosyan for his personal kindness and hospitality, and for being such a wonderful Director. Special thanks also to Tatyana Donskova, Danielle Métral and Egil Lillestøl for all their hard work before, during and after the School, which ensured that everything went smoothly and that everyone gained the maximum profit from the physics, and the location.

REFERENCES

[1] I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics*, 3rd ed., Vol. 1, (Institute of Physics, Bristol, 2003).

[2] I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics*, 3rd ed., Vol. 2, (Institute of Physics, Bristol, 2004).

[3] J. Goldstone, Nuov. Cim. **19** (1961) 154.

[4] Y. Nambu, in *New Theories in Physics*, Proc. XI Int. Symp. on Elem. Part. Phys., Kazimierz, Poland, eds Z. Ajduk, S. Pokorski and A. Trauman, (World Scientific, Singapore, 1989).

[5] V.A. Miransky, M. Tanabashi and K. Yamawaki, Mod. Phys. Lett. **A4** (1989) 1043.

[6] V.A. Miransky, M. Tanabashi and K. Yamawaki, Phys. Lett. **B 221** (1989) 177.

[7] W.A. Bardeen, C. Hill and M. Linder, Phys. Rev. **D 41** (1990) 1647.

[8] E. Farhi and L. Susskind, Phys. Rep. **74 C** (1981) 277.

[9] S.L. Glashow, Nucl. Phys. **22** (1961) 579.

[10] S.L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. **D 2** (1970) 1285.

[11] S. Weinberg, Phys. Rev. **D 8** (1973) 605 (especially footnote 8).

[12] G. Feinberg, S. Weinberg and P. Kabir, Phys. Rev. Lett. **3** (1959) 527 (especially footnote 9).

[13] C. Jarlskog, in *Fundamental Forces*, Proc. XXVII Scottish Universities Summer School in Physics 1 (1985).

[14] N. Cabibbo, Phys. Rev. Lett. **10** (1963) 531.

[15] M. Kobayashi and K. Maskawa, Prog. Theor. Phys. **49** (1973) 652.

[16] M.W. Grünewald, Phys. Rep. **322** (1999) 125.

[17] M. Veltman, Nucl. Phys. **B 7** (1968) 637.

[18] M. Velyman, Nucl. Phys. **B 21** (1970) 288.

[19] M. Veltman, Nucl. Phys. **B 123** (1977) 89.

[20] M. Chanowitz, E. Furman and I. Hinchliffe, Phys. Lett. **B 78** (1978) 285.

[21] M. Chanowitz, E. Furman and I. Hinchliffe, Nucl. Phys. **B 153** (1979) 402.

[22] J.J. van der Bij and M. Veltman, Nucl. Phys. **B 231** (1984) 205.

[23] J.J. van der Bij, Nucl. Phys. **B 248** (1984) 141.

[24] M.B. Einhorn and J. Wudka, Phys. Rev. **D 39** (1989) 2758.

GALACTIC AND SOLAR COSMIC RAYS

A. Chilingarian
Cosmic Ray Division, Yerevan Physics Institute, Alikhanian Brothers 2, Yerevan 36, Armenia
e-mail: chili@crdlx5.yerphi.am

Abstract
The aspects of Cosmic Ray (CR) origin are reviewed. Recent observational evidence on the spatial patterns of non-thermal X-ray radiation from Supernovae Remnants (SNR) supports long-awaited expectations of proton and nuclei acceleration up to PeV energies. We add new arguments based on the experimental data from surface arrays measuring Extensive Air Showers (EAS) and on data from solar accelerators available now from space-borne X-ray and gamma-ray spectrometers. Energy spectra of primary nuclei with atomic number from $Z = 1$ to $Z = 26$ can provide useful information on the validity of models of cosmic ray acceleration. By estimating the threshold energy of the onset of the suppression of the different nuclei flux, the so-called spectral 'knee' energy, we can directly check the hypothesis of rigidity-dependent acceleration of the hadrons in SNR sites. Unfortunately, information from the EAS experiments does not provide enough clues for such 'spectroscopy' of the 'knee region'. Nonetheless, by grouping the primary nuclei in two or three broad mass groups (light, intermediate and heavy) we can obtain useful information on energy spectra of the primaries. Recently, using multidimensional classification methods on MAKET-ANI experimental data, we categorized the 'all-particle' spectra into two distinct primary mass groups. From the spectra analysis, we come to the conclusion that the SNR-based particle acceleration model is valid and presents evidence that there exists a nearby source of cosmic rays, which provides a significant portion of the CR flux.

1 INTRODUCTION

The Cosmic Ray (CR) flux incident on the terrestrial atmosphere consists mostly of protons and heavier stripped nuclei accelerated at numerous galactic and extragalactic sites. The most exciting question associated with cosmic rays is the exploration of a particular astrophysical accelerating source. Owing to the bending in galactic magnetic fields, charged particles lose information about the parent sites during long travel and arrive on Earth highly isotropic. Only stable neutral particles i.e. X-ray, gamma quanta and neutrinos travel directly from sources and reveal exotic celestial objects and violent processes of their production. Orbiting telescopes and spectrometers, as well as ground-based Atmospheric Cherenkov Telescopes (ACT) and neutrino detectors have opened new windows to the Universe, detecting, in unprecedented detail, the spread of heavy elements during supernovae explosions, the ejection of the relativistic jets from black holes, and many other phenomena described in the last century only in science fiction.

A new paradigm in astrophysics research consists in the detection of celestial objects in radio, optical, X-, and gamma rays**.** A variety of compatible measurements gives sufficient information for building realistic models of physical processes of supernovae explosions, of accompanying gamma-ray bursts, of accretion disc interactions with super-dense objects, and finally of the evolution of the Universe itself. In this case additional information about the particles of highest energies will significantly enlarge the information on the most violent processes in the Universe and on the processes of the largest particle accelerators in space.

Galactic cosmic rays cannot map the objects where they are born, therefore, only integrated information from all sources is available from measurements of cosmic-ray fluxes near Earth and on the Earth's surface. This information consists of the shape of the energy spectra of the cosmic rays, their mass composition and their energy dependence, and of the anisotropy of the CR arrival.

Space-borne spectrometers on the ACE satellite, the AMS detector on the Space Shuttle, as well as numerous balloon-borne detectors measure the fluxes of different isotopes up to energies of 10 TeV rather precisely. Particle fluxes follow an overall power-law of $I(E)\ \alpha\ E^{\ \gamma}$ with spectral index of $\gamma \sim -2.7$. Therefore, because of the very low fluxes of the highest energy CRs and owing to very strict restrictions on the weight of the spacecraft payload, it is extremely difficult to get reliable information on particle fluxes above 10 TeV from space-borne spectrometers and calorimeters. However, it should be noted that recent successes with the long-lasting, new-technology balloon flights give hope that precise information on particle spectra up to several hundreds of TeV will be available soon.

Recently, the so-called kinematical method [1] was proposed, using thin (about 10 g/cm^2) target and silicon coordinate and charge detectors to precisely detect the charge and emission angles of secondaries produced in an inelastic interaction of primary nuclei. The angular distribution of the particles produced in the target carries information about the energy of a primary particle. This technique does not require total release of the energy as in the case of the ionization calorimeter, and the instrument could be made very light in weight. A one-year flight of such a device on the Space Station will provide data up to several PeV with 0.2 units of charge resolution. Currently there is no funded space experiment in the PeV region and, at least in the current decade, data will only be accessible from the Extended Air Showers (EAS) initiated by the 'primary' ion triggering a particle generation chain reaction in the terrestrial atmosphere and detected with large ground-based particle detectors. A variety of physical processes during the travel of the relativistic cloud of 'secondary' particles to the Earth's surface gave rise to different experimental methods, aiming to reconstruct the particle type, trajectory, and energy.

Signatures of the primary particles are microwave radio signals, fluorescent light, Cherenkov light, electrons, muons, neutrons, and hadrons reaching the Earth's surface and muons detected deep underground. The intensity and correlation matrix of each combination of mentioned signals carry information on the primary particles, but owing to the highly indirect nature of the experimentation, only some very robust characteristics of cosmic-ray fluxes of PeV and higher energy primaries have been unambiguously established up to now. First of all there is all particle energy spectra, reconstructed from so-called size spectra measured by plastic or liquid scintillators (so-called particle density detectors), distributed on the Earth's surface. Assuming a definite shape of the EAS electron lateral distribution function, and measuring the density of electrons on some rectangular or circular grid of distributed density detectors, and using a standard minimization analysis technique, the overall number of electrons (shower size) can be determined. By measuring the time delay of the arrival of the shower particles, using a system of distributed 'fast timing detectors', the zenith and azimuth angles of the shower core can be calculated (a very good estimate of the primary particle angles of incidence on the terrestrial atmosphere).

The shower size is correlated with the particle energy, but also with several unknown parameters such as particle type and the height of the first interaction. The functional form of the size–energy dependence introduces additional uncertainty, because it is obtained from a particular model of strong interaction of protons and ions with atmospheric nuclei, and at PeV energies there is no accelerator data to check this model. Different approximations of models fitted with manmade accelerator data at lower energies give significantly different results at higher energies. Nevertheless, during the last 50 years some important characteristics of spectra were established during intensive measurements with EAS surface detectors. For the list of detectors and their operational characteristics, see Ref. [2]. The most striking feature of the spectra is the approximately constant power index in the whole examined energy range. The power index slightly changes from value $\gamma \sim -2.7$ to value $\gamma \sim -3.0$ at 3–4 PeV (the 'knee', also known as suppression of spectra) and it is another important and well-established feature of the EAS spectra. Some authors [3], [4] claim that this 'knee' is a feature of only the size spectra,

reflecting some peculiarities of the EAS propagation and interaction in the atmosphere, and the flux of the cosmic rays incident on the atmosphere can be described by a constant power index and that the CR origin is of extragalactic nature. In the paper by Stenkin [5], the 'knee' of the EAS spectra is treated as a consequence of the shower size reconstruction method only. He demonstrates that the difference between pure electromagnetic showers and those having survived hadron 'cores' can be the cause of the 'knee'. Another, very interesting approach is connected with the enigma of supernovae implosion and collapse. In Ref. [6] the cannonball model of the supernovae explosion [7] was proposed as a source of the cosmic rays. The blobs of plasma with mass the size of Earth are ejected from poles of supernovae at nearly the speed of light. The population of such plasmoids filling the Galactic halo is responsible for the acceleration of the major part of the hadronic cosmic rays with energies up to another feature of all the particle spectra, the so-called 'ankle' occurring at $E > 10^{17}$ eV.

In contrast to these theories, the 'standard' models of CR acceleration name the Supernovae remnants (SNR) as a major source of CR. The detected non-thermal radio emission from SNR, which led to the natural assumption of the presence of accelerated electrons, made SNR the main candidate engine for particle acceleration [8]. Recent very detailed CHANDRA measurements of the X-rays from SN1006 [9] imply a very large effective magnetic field of 100 μG in the Supernovae remnant. In Ref. [10], the authors conclude that such a large field could be generated only due to the nonlinear interactions of the accelerated protons and stripped heavier nuclei with self-generated Alfven waves in a strong shock. Therefore, the SN1006 data confirms the acceleration of the nuclear component at least up to several units of 10^{14} eV. Gamma-ray pulsars usually located near the SNR centre are another candidate for cosmic-ray acceleration [11]. As mentioned in Ref. [12] pulsar-accelerated cosmic rays are expected to have a very flat spectrum. Therefore, the impact of the nearest pulsar to energies higher than 10^{14} eV can be tremendous and can explain the fine structure of the energy spectrum, which may reflect acceleration of the specific groups of nuclei.

To investigate various scenarios of particle acceleration in SNR, we still have to use indirect information contained in CR spectra in the vicinity of Earth. As Galaxy magnetic fields cannot confine particles with such high energies, the extragalactic origin of the highest energy particles is widely accepted. The MAKET-ANI detector, located at Mount Aragats in Armenia, owing to its modest size, is effectively collected cores of EAS initiated by primaries with energies up to several units of 10^{16} eV, therefore, we shall constrain our analysis to the energy range from 5×10^{14} eV to 2–3×10^{16} eV—the so-called 'knee' region. Energy spectra of primary ions from Z = 1 to Z = 26 will provide valuable information on the validity of the Standard Model. Information from the EAS experiments does not provide enough clues for such 'spectroscopy' of the 'knee region'. Nevertheless, precise measurements of the electron and muon content, and implementation of the CORSIKA simulation code by the KASCADE experiment [13] as we have demonstrated in numerous papers (see for example Refs. [14], [15], [16]), allow the classification of primaries according to three classes: 'light', 'intermediate', and 'heavy'. Using the nonparametric multivariate methodology of data analysis [17], [18], references on development and application of methods contained in Ref. [19], we solve the problem of the event-by-event-analysis of EAS data [20] using Bayesian and neural network information technologies [21], [22], [23].

At each stage of the analysis we estimate the value of the information content of the variables used for EAS classification and energy estimation and restrict the complexity of the physical inference according to this value. The MAKET-ANI experiment is located at 3200 m above sea level on Mt. Aragats, in Armenia; the quality of reconstruction of the EAS size and shape are good enough and we can use them for the EAS classification shower size and shape parameters (the so-called shower age). The distinctive information contained in distributions of these two parameters allows us to classify the EAS with a high level of accuracy into two distinct groups: initiated by 'light' or 'heavy' nucleolus. In the KASCADE experiment [24], where the muon content of the EAS is measured in addition to shower electron size, we can classify showers into three categories adding also the 'intermediate' class.

2 COSMIC-RAY ACCELERATION IN SUPERNOVAE EXPLOSIONS AND PROPAGATION IN THE INTERSTELLAR MEDIUM

The Power of Cosmic Ray (PCR) sources should be more than $\sim 10^{41}$ erg/s to maintain the estimated cosmic-ray energy density in the Galaxy. This value was obtained by multiplying the CR energy density in the Galaxy $\rho_{CR} \sim 10^{12}$ erg/cm^3 by the Galaxy volume $VG \sim 10^{67}$ cm^3 and dividing by the particle mean escape time from the Galaxy $\tau_{\rm esc} \sim 10^{14}$ s. The frequency of the SN explosions in the Galaxy (about one in 20 years), and the kinetic energy of supernova ejecta (about 10^{52} erg), lead to the CR luminosity of the same order of magnitude as if we assume that a few tens of per cent of the ejecta kinetic energy is transformed into the CR energy.

The power law is rather satisfactory to describe the spectra from 10^{12} eV (far above solar modulation effects), up to several units of 10^{18} eV, where the Galaxy magnetic field of $\sim$3 μG cannot confine the particles anymore. At low energies up to 10^{14} eV the spectral indices of protons, carbon, oxygen, and iron are very close to each other and equal to ~ -2.7. The same index describes the spectra of all the particles from 5×10^{14} up to 3–4$\times10^{15}$ eV (the knee region), where the power index changes to ~ -3, and returns to the value of -2.7 at several units of 10^{17} eV (the so-called ankle region). The 'classical explanation' of the changing behaviour of the spectra consists of the existence of three distinct acceleration mechanisms: the first, usually connected with SNR shock acceleration, fades in the knee region; the second, due to unknown causes, is responsible for energies from the knee to the ankle region; and the third, due to extragalactic sources, after the ankle.

Numerous papers are devoted to SNR-based acceleration. The obtained values of the spectra at the source obey the power law with index of $\gamma_s \sim (-2.0$ to $-2.1)$. Models of particle acceleration in the SNR can be compared with observations only if we account for the diffusion and escape of CRs from the Galaxy. Usually, energy dependence of the escape time is also taken from the power law $\tau_{\rm esc}\alpha E^{-\chi}$, and the relation between the spectra of CR in the source and the detected spectra takes the form $E^{-\gamma}\alpha E^{-(\gamma_s+\chi)}$. Theoretical calculations of the diffusion coefficient are based on assumptions regarding the distribution of magnetic inhomogeneities in the Galaxy. There are two main distributions: the 'Kolmogorov Spectrum', giving χ = 0.33 and the 'Kraichnan spectrum', giving χ = 0.5. Measurements of the spectra of low-energy isotopes ('radioactive clocks') gives another value of χ = 0.6. This value seems to be in perfect accordance with the observed spectra of $\alpha E^{-2.7}$, but it addresses only the low-energy particle data available from satellite and balloon isotope spectrometers. Additional measurements of isotope spectra at higher energies are needed. Since we cannot resolve the 'all-particle' spectrum, attempts are made to at least estimate the trend of the changing 'mean mass'. The calculations of the average depth of the shower maximum are made using data measured by the fluorescence and Cherenkov detectors' signal on 'lightening' of mean mass just before the knee, and transition to heavies above the knee. This behaviour could be explained by the influence of one or several of the nearest SNR, giving additional surplus flux added to the smeared superposition of thousands of Galaxy SNRs.

If the knee feature is due only to numerous distant sources, the steepening of the spectra should be much smoother than detected. Attempts to find time-temporal coordinates of the SNR, which would explain the observed fine structure of the spectra, heavily depend on the adopted energy dependence on the diffusion coefficient. Authors of the recent estimates of the possible location of the Single Supernovae (SS) [25], proceeding from the 'anomalous' diffusion introduced in Ref. [26], derive the following constraints for location and age of the SS—300–350 pc from the Sun and 90–100 kyr old. They also adopted the energy dependence of the diffusion coefficient with χ = 0.5. Very Long Baseline interferometric measurements of the 100 kyr old pulsar PSR656+14 [27] locate the pulsar in the centre of the SNR called Monogem Ring at 300 pc distance from the Sun. Therefore it was logical to assume that the Monogem Ring, the shell of debris from a supernova explosion, was the remnant of the blast that created the pulsar [28].

3 THE MAKET-ANI EXPERIMENT

The MAKET-ANI surface array [29] consists of 92 particle density detectors formed from plastic scintillators with a thickness of 5 cm. Twenty-four of them have an area of 0.09 m^2 and 68 an area of 1 m^2. The central part consists of 73 scintillation detectors and is arranged in a rectangle of 85 $\times$ 65 m^2. Two peripheral points at a distance of 95 m and 65 m from the centre of the installation consist of 15 and 4 scintillators, respectively Fig. 1. In order to estimate the zenith and azimuthal angles, 19 detectors out of the 92 (each with an area of 1 m^2) are equipped with timing readouts to measure the timing of the appearance of the EAS front with an accuracy of $\sim$5 ns. The photomultiplier tubes (PM-49) of the detectors are placed in light-tight iron boxes. Logarithmic Analog to Digital Converters (ADCs) and Constant Fraction Discriminators (CFDs) are assembled just above the photomultiplier tube (PM). The dynamic range of the registered particle number is $\sim 5 \times 10^3$.

Two types of detector triggers are used:

1. The hardware trigger: at least 7 out of 11 central density detectors must be hit with more than 3 particles.
2. The timing trigger: at least 4 out of 9 timing detectors, symmetrically arranged relative to the centre, must be hit.

If the first two conditions are fulfilled in a time window of 20 μs, then the event is stored. The trigger and data readout systems are according to the CAMAC standard. Monte Carlo calculations show that this trigger system selects EAS with sizes $N_e > 5{\times}10^4$ and cores located within the rectangle of 40 $\times$ 12 m^2 around the geometrical centre of the installation.

The uncertainties of the reconstruction of the EAS parameters are as follows:

Shower size $\Delta\, N_e \sim 10\%$, the shower shape (age) parameter—$\Delta\, s \sim 0.06$.

The accuracies of the EAS angle determination are $\Delta\theta \sim 1.5^\circ$ and $\Delta\phi < 5^\circ$.

In the period from 1998 to 2002, approximately 7 788 000 EASs were registered with effective registration time of about 24 000 hours. From these showers only $\sim$ 963 000 events were selected for the spectra calculation. The selection criterion was to have more than 95% efficiency of registration, so we selected the EAS core from the more compact area around the geometrical centre of the MAKET detector, ensuring high efficiency of EAS registration. The following cuts were applied for the events selection:

$N_e > 10^5$, $0.3 < s < 1.7$, -24 m $< X_0 < 24$ m, -12 m $< Y_0 < 12$ m, $\theta < 45^\circ$.

During multiyear measurements, the detecting channels were continuously monitored. Data on background cosmic-ray spectra was collected for each detector. The slope of the spectra was used for detector calibration. The slope of background spectra is a very stable parameter and does not change even during very severe Forbush decreases, when the mean count rates can decrease as much as 20% [30]. The detailed information about the MAKET-ANI detector operation during 1997–2003, including various comparisons and uniformity checks are summarized in Refs. [31], [32].

4 SELECTION OF EAS PARAMETERS FOR CLASSIFICATION AND ESTIMATION

We are interested in choosing a combination of the EAS measured characteristics significantly differing from light and heavy initiated showers. The discriminative power of EAS characteristics was investigated using CORSIKA [13] and MAKET-ANI response simulation codes [31]. For comparison of EAS initiated by different primary ions a number of statistical methods were used, including one-dimensional statistical tests, correlation analysis, and misclassification rates estimation by neural and Bayesian classificators. Input parameters of the simulation program included particle type, energy, angles of incidence, as well as geographical coordinates and altitude of the MAKET-ANI detector. The energy and angular distributions taken reflect modern theoretical expectations. Owing to the stochastic nature of particle

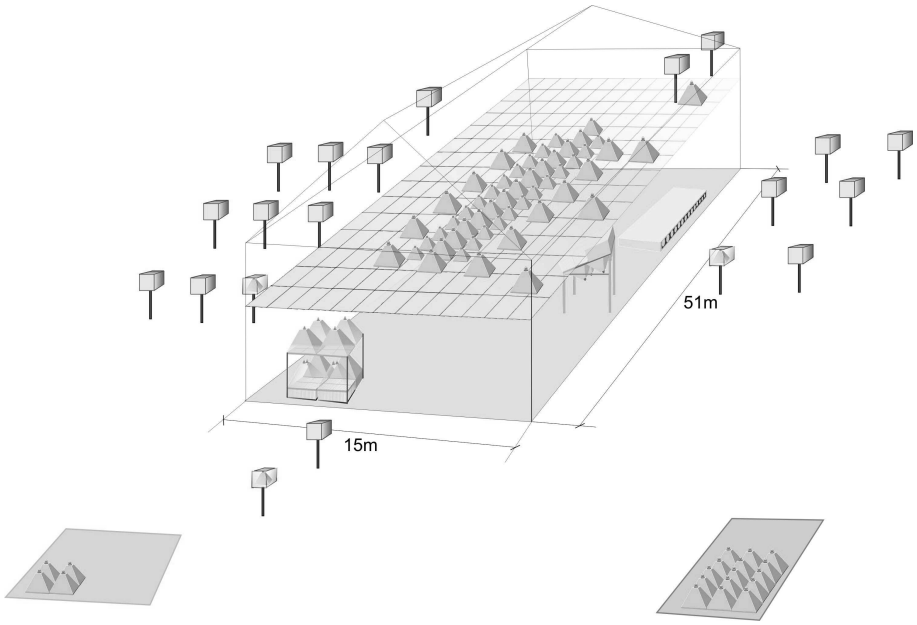


Fig. 1: MAKET-ANI detector setup

propagation through the atmosphere, the output parameters of simulation programs are random variables. We proceed according to the assumption of two-way division of all primary nucleolus, the so-called, 'light' and 'heavy' mass groups. As representatives of the light group we shall take the proton and He nucleus; for the heavy group the Si and Fe nuclei will be the representatives. The intrinsic differences of the light and heavy ion cascades in the atmosphere make the distributions of EAS parameters different. We investigate if this difference is sufficient for reliable two-way classification and take into consideration the way that the detector response smears it. Integrated over the entire energy range, the shower sizes of EAS initiated by heavy and light nuclei are also very similar. The only parameter showing a significant difference between the two is the shower shape-age (s) parameter. Although the detector smears this difference, it remains significant enough, and, as we shall see further, the various correlations of this feature with shower size make the pair of parameters (N_e, s) effective both for classification and energy estimation.

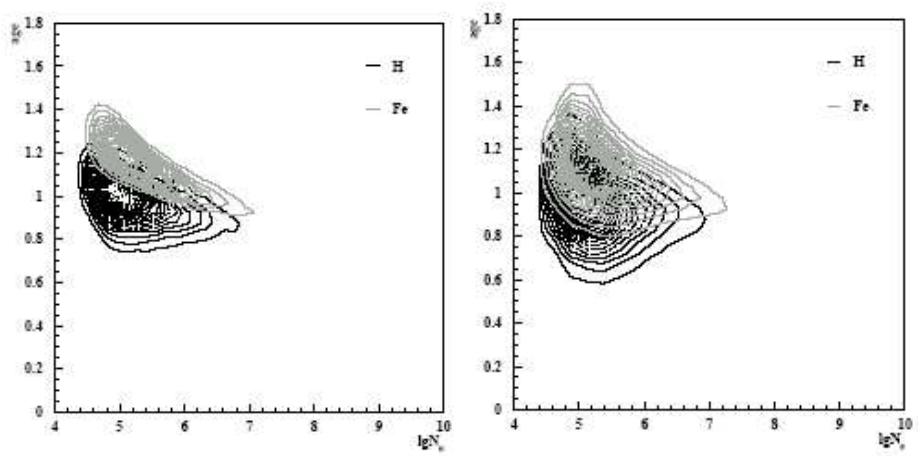


Fig. 2: Scatter plot of shower age versus shower size for simulated light and heavy primaries with (right) and without (left) incorporating the detector response

The most direct estimates of the 'discriminative power' of EAS characteristics are obtained by the classification of the samples using EAS simulations. Overlapping of the 2-dimensional distributions apparent from Fig. 2 could be calculated numerically by the estimation of the misclassification rates from Bayesian or neural network classification of EAS initiated from the alternative groups of nuclei. Using only EAS electron characteristics, we cannot resolve nuclei with similar masses, such as p and He, or Fe and Si, therefore we join these nuclei in groups naming them 'light' and 'heavy', thus restricting ourselves to the two-way classification of the experimental data. Expected classification results posted in Table 1 and Table 2 demonstrate that although detector smearing significantly enlarges misclassification rates, nevertheless >70% correct classification is very encouraging and the N_e–s pairs as measured by the MAKET-ANI detector provide enough information for the two-way classification. We also want to point out the good agreement between results obtained by using two completely different methods of classification: Bayesian classification with nonparametric estimation of multivariate probability density function and neural network classification using stochastic net training methodologies.

Table 1: Neural classification into two classes H+He and Si+Fe events without and with detector response

	Without detector response		With detector response	
	Light	Heavy	Light	Heavy
Light	0.925	0.075	0.720	0.280
Heavy	0.045	0.955	0.240	0.760

Table 2: Bayesian classification into two classes using H+He and Si+Fe events without and with detector response

	Without detector response		With detector response	
	Light	Heavy	Light	Heavy
Light	0.938	0.062	0.712	0.288
Heavy	0.043	0.957	0.237	0.763

5 DATA CLASSIFICATION INTO LIGHT AND HEAVY GROUPS OF NUCLEI, PURIFICATION OF SELECTED GROUPS OF NUCLEI

According to the results from the previous section we use two 'training samples' of 'light' and 'heavy' nuclei initiated N_e–s pairs, generated by the CORSIKA code including the MAKET-ANI response function. Before neural classification of the MAKET-ANI data we investigate the expected purity[1] and efficiency[2] of our data analysis procedures. From Table 3 we can see that efficiency of classification, i.e., correct identification of nuclei from light and heavy groups is above 70%, the 'intermediate' oxygen nuclei are distributed approximately equally among two groups. To obtain purity estimates we assume the so-called 'normal' primary composition: 30% H, 24% He, 17% O, 17.5% Si and 11.5% Fe.

Table 2 demonstrates that the purity of the light group is above 70% and the purity of the heavy group is below 50% with large contamination of the oxygen and light nuclei.

To enlarge the purity of the heavy nuclei group we introduce the purification procedure described in Ref. [16], enlarging the purity of each nuclear group at the cost of decreasing the efficiency. The purification of the selected 'light' and 'heavy' groups was done by selecting the appropriate domain in the entire range of the network output. The feed-forward Neural Network (NN) performs a nonlinear mapping of the multidimensional characteristics of the EAS to the real number interval [0,1], called the

[1]Purity: fraction of true classified events in an actual number of events assigned to a given class.
[2]Efficiency: fraction of true classified events in total number of events of a given class.

Table 3: Efficiency of the neural classification of EAS initiated by different primaries into two mass groups

	Light	Heavy
H	0.720	0.280
He	0.691	0.309
O	0.453	0.547
Si	0.352	0.648
Fe	0.240	0.760

Table 4: Purity of the classification of different nuclei in light and heavy groups

	H	He	O	Si	Fe
Light	0.407	0.298	0.137	0.111	0.047
Heavy	0.162	0.167	0.208	0.255	0.208

output of the NN. Figure 3 shows the network output histogram. The network was trained to shift the 'heavy' group to the right and the 'light' group to the left of the histogram. The 0.5 point of the NN output is the so-called decision point. The particular class assignments for the two-way classification are the subintervals [(0.0,0.5) and (0.5,1.0)] for the 'light' and 'heavy' class, respectively. If the neural network is satisfactorily trained to have generalization capabilities, the output distributions for the different classes will overlap at the subinterval boundaries. Therefore, by shrinking the subintervals, i.e. moving the interval boundary to the left and rigsht of the decision point 0.5, it is possible to remove a large portion of the misclassified events. Of course, simultaneously we lose parts of the true-classified events, i.e., decrease the efficiency. Thus, instead of one decision point in the middle of the NN output interval, we have two 'decision intervals' for accepting 'light' and 'heavy' nuclei, and a third interval in between where we reject the classification. Figure 3 demonstrates this 'purification' procedure.

Figure 4 shows the results of the purification. The values next to the symbols indicate the selected decision interval used for obtaining the particular purity–efficiency relation. For example, if we select the [(0.0,0.3) and (0.7,1.0)] intervals for classification of the 'light' and 'heavy' nuclei, we obtain 96% purity and 56% efficiency for the 'light' class; 78% purity and 55% efficiency for the 'heavy' class. Therefore, we can enhance the purity of the light nuclei up to 95% and the purity of the heavy nuclei up to 80%, while still holding the efficiency above 50%. The purity and the efficiencies are obtained by classifying 35 000 light (H,He) and 17 000 heavy (Si,Fe) control events, which are not used for the training of the neural network. Artificially high purity for both classes is achieved by using this method as demonstrated in Table 4, since the intermediate nuclei (simulated oxygen initiated EAS) were not included in the analysis. More realistic purity and efficiency estimates are apparent from Table 5 and Table 6, where we include also the oxygen nuclei.

As we can see from Table 6 the purity of the light group increases from 70% to 77% and for the heavy ones from 46% to 55%; we need to keep in mind that approximately 20% of the heavy group are due to showers initiated by oxygen nuclei. The purification allows us to significantly increase the purity of two alternative samples and we can, therefore, estimate the energy spectra of light and heavy groups. Of course, first we should describe the energy estimation procedures used.

6 ESTIMATION OF THE PRIMARY ENERGY OF DIFFERENT GROUPS OF NUCLEI

The primary energy of each shower was obtained by neural network estimators separately for the light and heavy nuclei induced events, exploiting the very large correlation of shower size N_e with primary

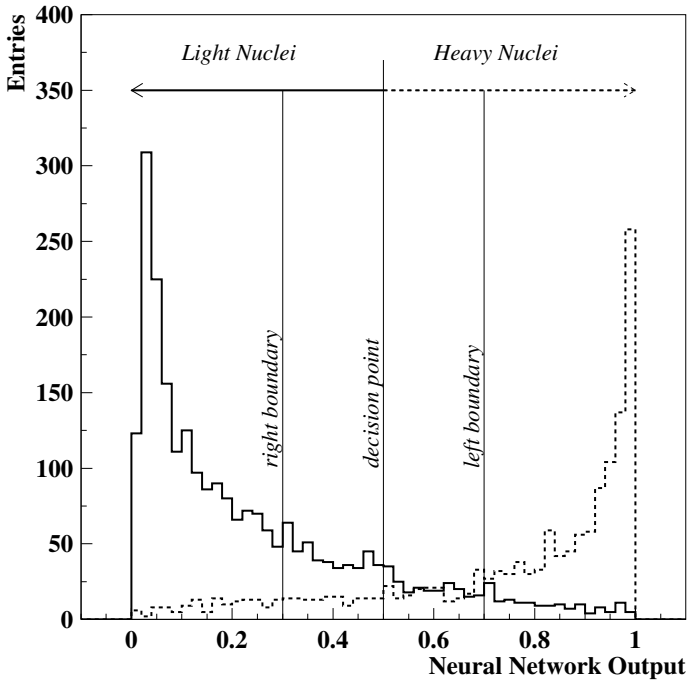


Fig. 3: Output of the Neural Network (NN) trained to distinguish light and heavy nuclei

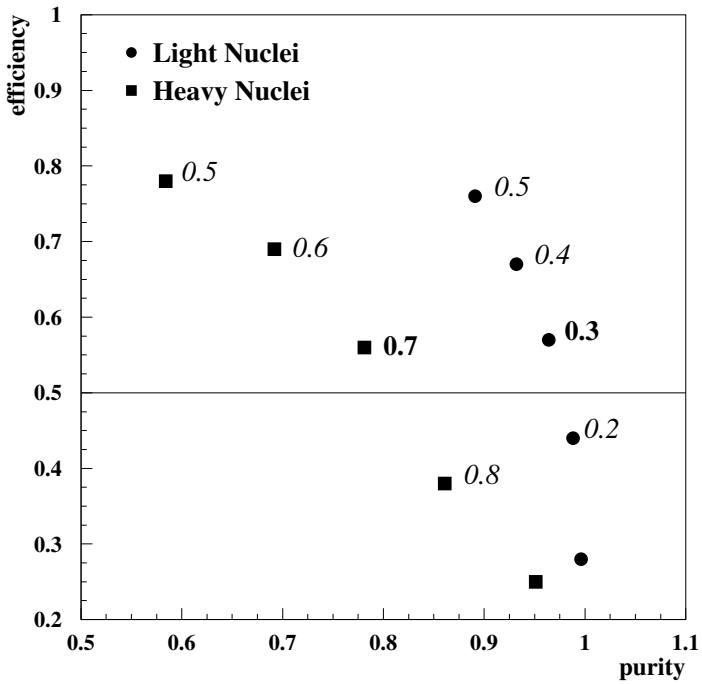


Fig. 4: Purity–efficiency plots obtained by shifting the NN decision boundaries

Table 5: Efficiency of the neural classification of EAS initiated by different primaries into two mass groups (purification intervals [(0.0,0.3) and (0.7,1.0)])

	Light	Heavy
H	0.567	0.095
He	0.475	0.135
O	0.252	0.303
Si	0.176	0.393
Fe	0.099	0.561

Table 6: Purity of the classification of different nuclei in light and heavy mass groups (purification [(0.0,0.3) and (0.7,1.0)])

	H	He	O	Si	Fe
Light	0.459	0.310	0.115	0.084	0.032
Heavy	0.115	0.131	0.207	0.278	0.268

energy and different correlations between primary energy and shower shape in light and heavy nuclei groups. In Fig. 5 relative errors of energy estimation for 10 energy intervals are posted. The bias of the energy estimation displayed does not exceed 5% for the light group (left) in the whole energy range except the lowest energies. For the heavy group of nuclei the estimation bias in the energy range of 10^{15}–10^{16} eV is not larger than 5%, nevertheless, one can observe some overestimation for low and high energy regions. The energy resolution for the heavy group of nuclei is significantly better (MSD $\sim$20%) as compared to the light group of nuclei (MSD $\sim$30%) due to the smaller fluctuations of heavy initiated EAS size and shape. Also, the accuracy of the energy estimation is enhanced with enlarging primary energy.

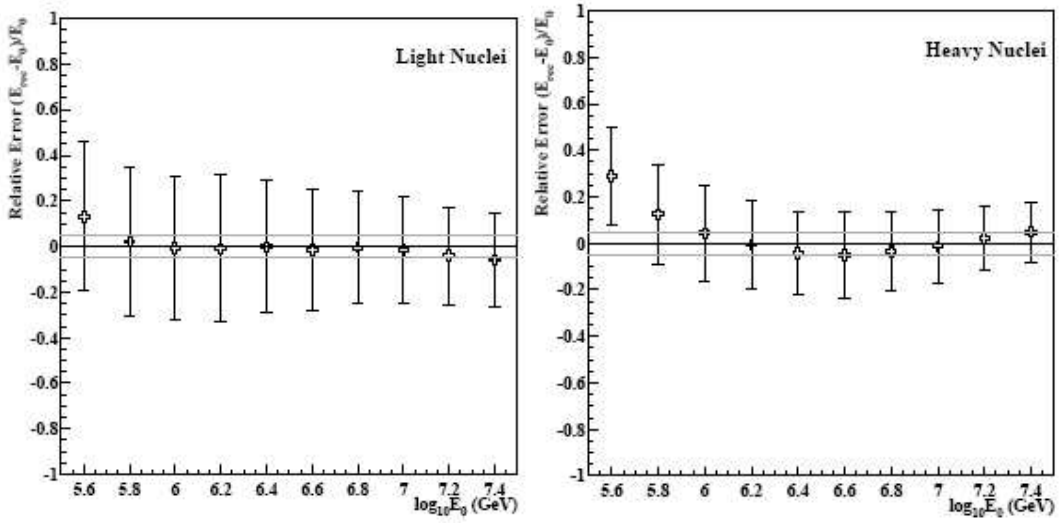


Fig. 5: The relative errors of energy estimates for 10 energy intervals of light and heavy groups, the two horizontal lines around the 0-line outline the $\pm$5% error corridor. Error bars correspond to root mean square (r.m.s.) deviation

7 ENERGY SPECTRA

Figure 6, adopted from Ref. [2], shows the energy spectrum measured by different detectors exploiting various experimental techniques and energy reconstruction methods. Energy estimation for all experiments was made using Monte Carlo simulations with different numerical algorithms. Despite considerable differences in experimental techniques and different EAS components (shower shape and electron size parameters, muons, Cherenkov light) used for the energy estimation, and the differences in systematic errors (usually not reported in publications), almost all spectra are in rather good agreement if we assume an energy estimation accuracy of $\sim$20%. Only at energies higher than the knee feature do the spectra disagree, probably because of the saturation effects in the scintillators in some experiments. All particle spectra and mean logarithmic mass, in many cases presented as an outcome of the EAS experiment, are not very informative. We never know which combination of primaries constitutes the mean and which groups of primaries are responsible for the knee. The best solution will be to separate different groups of nuclei and reconstruct energy spectra to determine the spectral knees of different nuclei at different positions. This programme was partly fulfilled with the data from the MAKET-ANI experiment. After checking for the purity and the efficiency of each of the nearly 1 million showers registered by the MAKET-ANI installation in 1999–2002, shower sizes greater than 10^5 were classified according to the techniques described in Refs. [20], [16]. The energy of the classified particles in two distinct classes of showers was estimated for each group separately, again using the CORSIKA simulations and neural estimation techniques. Using the EAS characteristics of shower size (N_e) and shape (s), we plot the obtained energy spectra of the 'light' and 'heavy' mass groups, see Fig. 7.

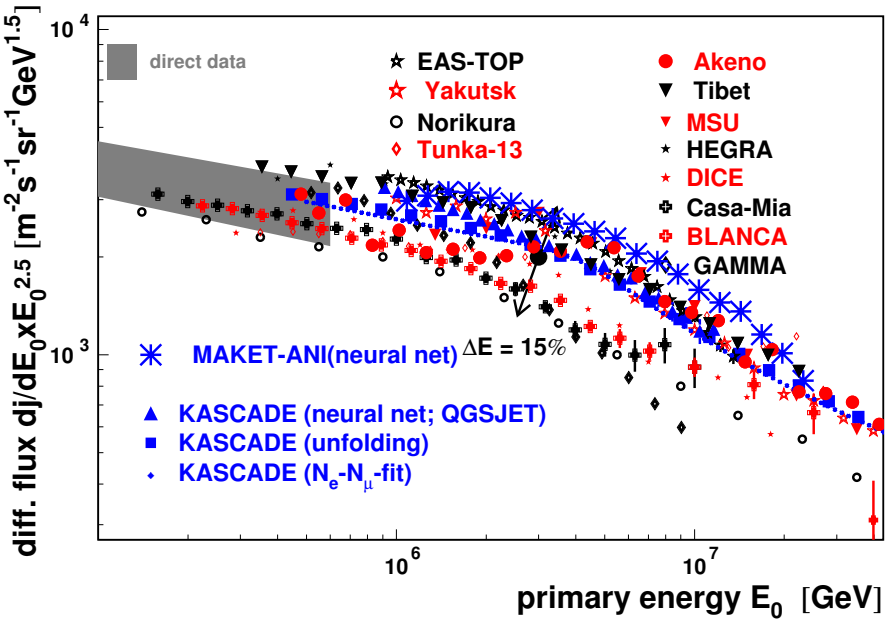


Fig. 6: Summary of the all particle spectra from 18 different experiments

The spectrum of the 'light' group shows a 'knee' in the region of 3–4$\times 10^{15}$ eV. The 'knee' feature is not observed for the spectrum of the 'heavy' component, at least not up to energies of 10^{16} eV. The number of 'light' and 'heavy' nuclei at $\sim 10^{15}$ eV is approximately equal and the number of 'heavy' nuclei gets larger at energies greater than the 'knee' energy. The 'purified' spectra, see Fig. 8, show lower flux intensities for both classes of particles due to the lower efficiency. The 'knee' position shifts to lower energies because, after purification, the proportion of protons is enlarged. In addition, the slope of the spectrum (spectral index) of the 'purified' light component becomes steeper, -2.63, compared to = -2.54 before purification. Both results are consistent with the rigidity-dependent acceleration and consequent fading of the proton flux at high energies.

Another important feature of the obtained spectra is the very large difference between spectral indices before and after the 'knee': $\Delta\gamma$(light) = $\gamma_2 - \gamma_1 \sim 0.9$. It is well known that the same parameter for the all-particle spectra is $\Delta\gamma$ (all-particle) ~ 0.4, [2]. Erlykin and Wolfendale, in their simulations, were not able to reproduce the actual shape of the all-particle spectrum by averaging the proton and nuclei fluxes produced by nearly 50 000 distant supernovae in our Galaxy [33]. Therefore, they propose that the nearby young supernova (< 500 pc and < 110 kyr), is responsible for the approximately 60% of the detected cosmic-ray flux in the vicinity of Earth [25]. The very large difference of the spectral indices before and after the knee of the 'light' component (~ 0.9) confirms the Erlykin and Wolfendale proposal regarding the huge impact of the nearest supernova on the cosmic-ray flux in the vicinity of Earth. It suggests the necessity to make detailed calculations of the influence of the nearest supernova on the detected cosmic-ray fluxes, i.e., to obtain the partial spectra of the nuclei accelerated by the single source (for a candidate of such a source see Ref. [28]).

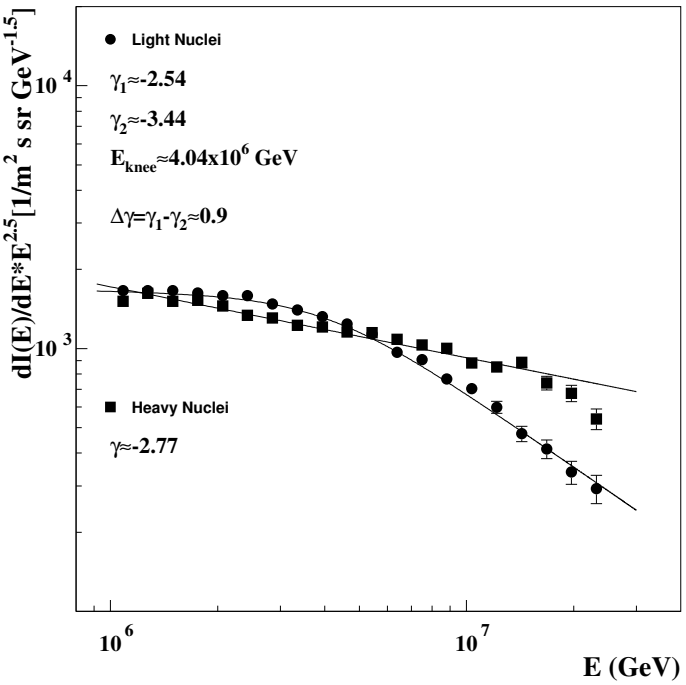


Fig. 7: Energy spectra of light and heavy nuclei obtained by neural classification and energy estimation. EAS characteristics used: shower size and shape (age)

8 WHAT WE CAN LEARN FROM SOLAR ACCELERATORS

With the launch of particle spectrometers in the 1970s, began the continuous monitoring of low- and medium-energy cosmic rays in space. Time histories of the simultaneously detected X-rays, gamma-rays, electrons, and ions of different energy and charge, combined with the detection of the developing flares and Coronal Mass Ejections (CME) using coronagraphs, helped to create a comprehensive picture of the major solar events, accelerating protons to high energies, the so-called, Solar Energetic Proton (SEP) events [34]. SEP events include also highest energy ions and accompanying protons, giving rise to Ground Level Enhancements (GLE) and additional fluxes of secondary cosmic rays (mostly neutrons and muons), detected by the world-wide network of Neutron Monitors and Muon Telescopes. "New Instruments on WIND and ACE satellites operating during the 23rd solar cycle, with geometry factors $\sim$100 times larger than those of the previous cycle, have yielded unprecedented observations of

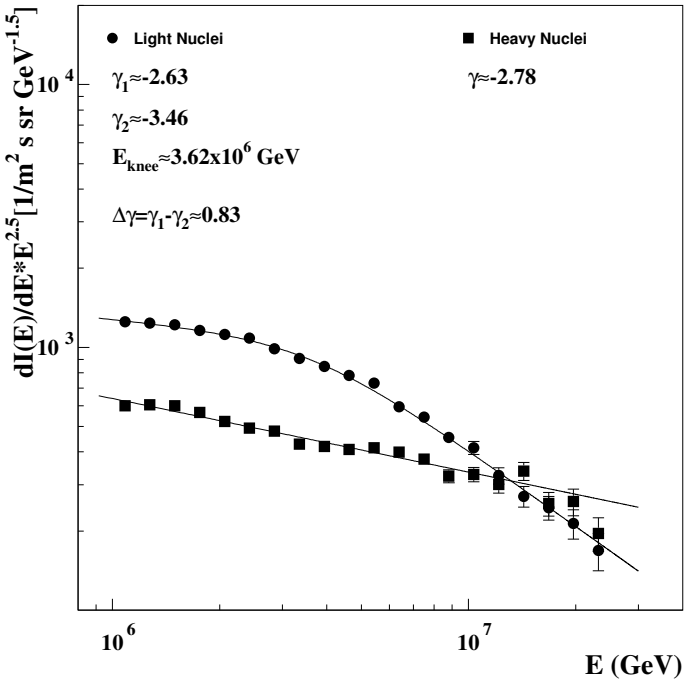


Fig. 8: Energy spectra of light and heavy nuclei obtained by neutral classification and energy estimation. The same as in Fig. 7 but obtained with purified light and heavy data samples. Purification intervals: [0.0,0.3) and (0.7,1.]

temporal evolution in composition and spectra over a wide range of energies and species" [35]. Multiwavelength measurements from very sensitive X-ray detectors, high-resolution imaging coronagraphs and radiotelescopes now reveal the location and characteristics of the natural accelerators at the Sun and in interplanetary space in much more detail.

Impulsive flare events are believed to accelerate electrons and ions in large structures originating in the magnetic reconfiguration regions. After discovery of the above-the-loop-top hard X-ray source [36], with the Yohkoh/HXT [37], it became apparent that particles are accelerated by the dynamic electromagnetic forces during the reconfiguration of the magnetic fields [38]. The most probable acceleration mechanism is stochastic acceleration, allowing detectable intensities of nonthermal X-ray radiation from locally trapped electrons. Direct hard X-ray detection, as well as application of the time-of-flight technique to the electrons travelling from the acceleration site to the chromosphere, reveals that the location of the acceleration region is 5000–35 000 km above the top of the soft X-ray-bright flare loop [39]. The natural assumption that positively charged protons and ions will be accelerated with the same mechanisms as the electrons is proven by the registration of the lined gamma radiation in coherence with hard X-ray radiation. The time sequence of the bremsstrahlung radiation peaks produced by accelerated electron beams, interlaced by the nuclear de-excitation lines produced by proton and ion bombarded chromosphere, clearly demonstrates that ions and electrons are accelerated in the same region and nearly simultaneously. The efficiency of the stochastic acceleration of ions via the mutual wave-particle interactions depends on the relation between the frequencies of the resonant waves (Alfven waves, magnetosonic waves, sound waves) and ion gyrofrequency. Alfven waves, if fast enough ($\sim$2000 km/s), can accelerate 20 keV protons up to GeV energies during time scales of 1–10 s [40], [41].

Gradual events are associated with CME development in corona and in interplanetary space. CME driven shock should be fast enough ($>$ 500 km/s, [34]) to produce SEP events. Shock acceleration is believed to be one of the major mechanisms in the Universe for accelerating particles to highest en-

ergies. Multiple traversals of shock are required for the acceleration of solar ions up to MeV energies. Ambient magnetic turbulence is not sufficient for scattering and trapping ions with such energies. Self-generated Alfven waves effectively scatter energetic ions, providing their trapping near the shock and, therefore, increasing their energy. The maximum attainable energy of accelerated ions is proportional to the rate of re-crosses of the shock. This rate, in turn, is proportional to the particle trapping time. "As trapping increases for particles of one rigidity, they are more likely to be accelerated to a higher rigidity, where they again stream out and produce resonant waves, etc." [42]. Numerical calculations and Monte Carlo simulation prove that solar protons could be accelerated up to energies of 100 GeV during propagation of the CME in middle and high corona [41]. The same authors, examining the 1982 June 3 flare, mention that protons were accelerated within 16 seconds from 30 MeV to ∼1 GeV. Krucker and Lin [43], based on the data from WIND/SST instrument [44], conclude that protons at energies up to 6 MeV are injected simultaneously at heights $\leq 10\ R_{\oplus}$. The maximum energy attainable by the shock acceleration depends on the shock speed and the height of the shock starts in the corona. Shock waves as fast as ∼1500 km s^{-1} starting below $\sim\ 5\ R_{\oplus}$ can accelerate ions up to 10–30 GeV [45], [46]. Study of the association between SEP events and CME [47] proves that CME interaction is important for high-energy SEP production. For most of the SEP events detected in 1997–2001 the primary, fast CME overtakes one or more slower CMEs within a heliocentric distance of $\sim\ 20\ R_{\oplus}$. The summary of the present knowledge on particle acceleration by various mechanisms at the Sun and in interplanetary space is as follows:

- Electron accelerators also accelerate protons and heavier ions, acceleration sites are very close in space and time.
- Particle acceleration is much more effective when several shocks are present in interplanetary space.
- The 'magnetic bottle' structures formed by interacting shocks are major sites for reacceleration of particles primarily accelerated by 'impulse' and 'gradual' mechanisms.
- Maximal attainable energy of particle accelerators is proportional to the particle charge.
- Moving shock carries the bulk of particles.
- The maximal attainable energy of the particle changes from event to event and depends on the total energy of the solar blast, speed of the shock wave, and the time-temporal history of the solar flare. (Positions of the Spectral 'knees' change from 10 MeV to several GeV.)
- For detection of charged particles on the Earth, the shock should intercept the observer's magnetic tube.
- The streaming limit controls the transport of particles.

The rigidity-dependent maximal acceleration energy in Solar Energetic Events (SEP), which occurred during the current 23rd solar cycle, is apparent from Fig. 9. Again, as with galactic cosmic rays, we see a very sharp knee for the light nuclei group, namely protons, and no knee for the heavy nuclei group, namely iron.

The most famous, so-called Bastille day SEP from the 14 July 2000 event, as you can see from Fig. 9 and from Ref. [46] demonstrates the remarkable exactness of the knee positions according to accelerated ion charge: the proton knee is at ∼20 MeV, the helium knee at ∼40 MeV, and the carbon knee at ∼100 MeV. The carbon charge is equal to 5, and one should note that the temperature at and near the Sun is not high enough to fully strip the carbon ions, such as happens at Supernova explosion sites.

9 DISCUSSION AND CONCLUSIONS

Recent unprecedented detailed observations of the nonthermal X-ray radiation from SN1006 made by CHANDRA [9], point very definitely at the SNR as the host of the hadron accelerators providing energy at least up to several units of 10^{14} eV. Observations of the wind synchrotron nebulae around pulsars in the vicinity of the SNR centre reveal another accelerator site, e.g., the termination shock, at which the relativistic shock from the pulsar wind is forced to join the slower expansion of the outer nebula [48].

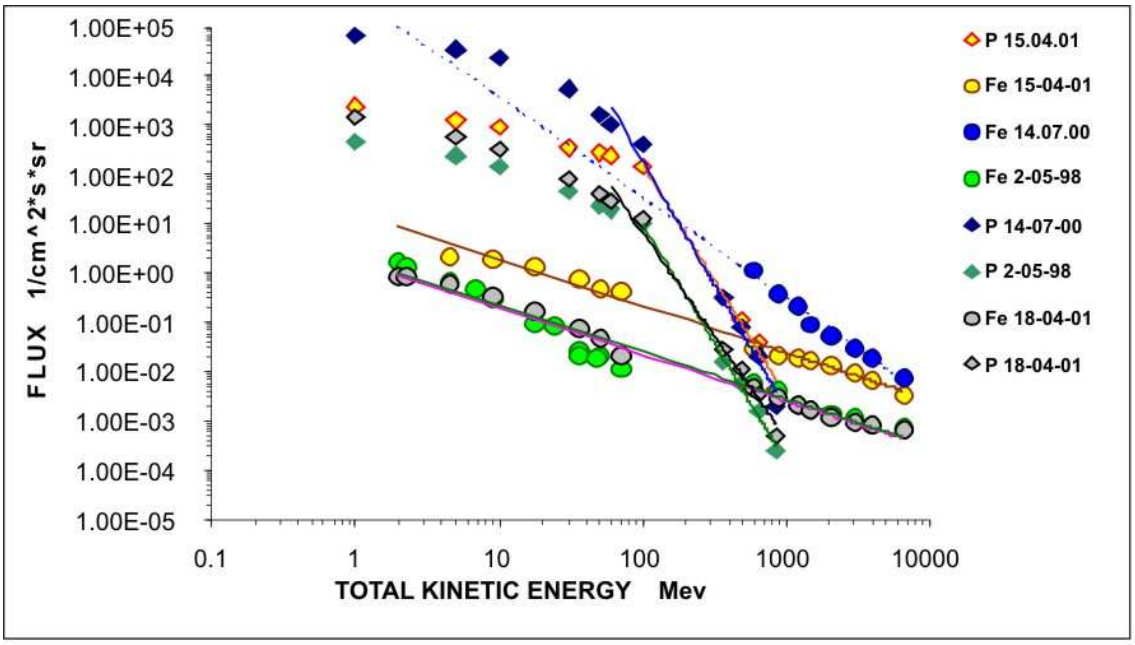


Fig. 9: The energy spectra of protons and Fe ions registered by space-born detectors during an SEP event of 23-rd solar cycle

The Single Source, or Single Supernovae (SS) model of Erlykin and Wolfendale [49], [50], also attracted huge support by the discovery of the nearest pulsar, located at the space-temporal distance in remarkable concordance with SS model expectations [28].

The recent results, which came forth from the MAKET-ANI experiment, confirm the SNR and SS models of cosmic-ray origin. The very sharp knee of the energy spectrum of the light mass group suggests accepting the SS hypothesis, because it is highly improbable that a Galaxy ensemble of distant supernovae, with a variety of explosion energies, shock-wave speeds, distances and explosion times, will provide a sharp knee feature. Instead we should expect rather smooth depletion of the light mass group flux if the latter hypothesis were true.

The knee of the light mass group and the absence of knee in the heavy mass group up to at least 10 PeV also supports the hypothesis of rigidity-dependent maximal energy of SNR accelerators.

Experimental evidence could be summarized in the following statements:

– The energy spectrum of the ‘heavy’ mass group of cosmic rays shows no ‘knee’ in the energy interval of 10^{15}–10^{16} eV.
– The energy spectrum of the ‘light’ mass group of nuclei shows a very sharp ‘knee’ $\Delta\gamma \sim 0.9$ compared to $\Delta\gamma \sim 0.4$ for the all-particles energy spectra.

And finally we conclude that:

– The SNR acceleration model is supported by the MAKET-ANI data on partial energy spectra.
– Our conclusions on rigidity-dependent acceleration are consistent with the evidence we saw recently about how the solar accelerators work. The recently proposed mechanisms of particle acceleration in the SN1006 [10] is fully consistent with the mechanisms of solar particle acceleration by CME driven shocks, of course, at much lower particle energy scales.
– The time history of the cosmic-ray intensity [51] suggests 50% enhancement of the CR flux integrated over the last 400 000 years, compared to an all available time record of 10^9 years. It is also consistent with the nearby SS model.

ACKNOWLEDGEMENTS

We thank the ANI Collaboration members for their fruitful cooperation over many years and A. Haungs, H. Rebel and M. Roth for cooperation in development of multivariate methods of data analysis for the KASCADE experiment and for kind permission to use Fig. 6.

REFERENCES

[1] J. Adams, G. Bashindzhagyan et al., Adv. Space Res. **27**, No. 4, (2001) 829–833.

[2] A. Haungs, H. Rebel and M. Roth, Rep. Prog. Phys. **66** (2003) 1145–1206.

[3] S.I. Nikolsky and V.A. Romakhin, Phys. Atom. Nucl. **63**(10), (2000) 1799.

[4] S.I. Nikolsky and V.G. Sinitsyna, Proc. 28th ICRC, Tsukuba, (2003) pp. 2007–2010.

[5] Y. Stenkin, Mod. Phys. Lett. A, **18**(18), (2003) 1225–1234.

[6] R. Plaga, New Astron. **7** (2002) 317.

[7] A. Dar and R. Plaga, Astron. Astrophys. **349** (1999) 259.

[8] K. Koyama, R. Petre et al., Nature **378** (1995) 1306.

[9] K.S. Long, S.P. Reynolds, J.C. Raymond et al., Astrophys. J. **586** (2003) 1162.

[10] E.G. Berezhko, L.T. Ksenofontov and H.J. Volk, Astron. Astrophys. **412** (2003) 11.

[11] W. Bednarek and R.J. Protheroe, Astropart. Phys. **16** (2002) 397.

[12] A. Bhadra, Proc. 28th ICRC, Tsukuba, (2003) p. 303.

[13] D. Heck et al., FZKA 6019, Forschungszentrum Karlsruhe (1998).

[14] A.A. Chilingarian et al., for the KASCADE Collaboration, (1997), In Proceedings of 28th ICRC, Durban, v. 4, (1997) 105–108.

[15] A.A. Chilingarian, M. Roth and A. Vardanyan, for the KASCADE Collaboration, J. Phys. G, **75A** (1999) 302.

[16] T. Antoni et al. (KASCADE Collaboration), Astropart. Phys. **19** (2003) 715–728.

[17] A.A. Chilingarian, Comput. Phys. Commun. **54** (1989) 31.

[18] A.A. Chilingarian, *Analysis and Nonparametric Inference in High Energy Physics and Astroparticle Physics*, Program Package ANI, (Users Manual, unpublished), (1998), http://crdlx5.yerphi.am/proj/ani

[19] A.A. Chilingarian and A.A. Vardanyan, Nucl. Instrum. Methods, **502**/2-3 (2003) 787.

[20] A.A. Chilingarian and H.Z. Zazyan, Nuovo Cim. **14C**(6), (1991) 555.

[21] A.A. Chilingaryan, Neurocomput. **6** (1994) 497.

[22] A.A. Chilingarian, Patt. Recogn. Lett. **16** (1995) 333–338.

[23] C.M. Bishop, *Neural Networks for Pattern Recognition* (Oxford Univ. Press, New York, 1995).

[24] T. Antoni et al. (KASCADE Collaboration), submitted to NIMA, (2003).

[25] A.D. Erlykin and E.W. Wolfendale, Proc. 28th ICRC, Tsukuba, (2003) pp. 2349–2052.

[26] A.A. Lagutin, Proc. 27th ICRC, Hamburg, **5** (2001) 1900.

[27] W.F. Brisken, S.E. Thorsett, A. Golden and W.M. Goss, Astrophys. J. **593** (2003) L89.

[28] S.E. Thorsett, R.A. Benjamin, W.F. Brisken, A. Golden and W.M. Goss, Astrophys. J. **592** (2003) L71.

[29] A. Chilingarian, G. Gharagyozyan et al., Proc. of the Workshop ANI 99, Eds. A. Chilingarian, A. Haungs, H. Rebel, M. Zazyan, Nor Amberd, 1999, FZK preprint 6472 (1999).

[30] A.A. Chilingarian and V.Kh. Babayan et al., Adv. Space Res. **31** (2003) 861–5.

[31] G.G. Hovsepyan, ANI Collaboration report No. 2 (2002) http://crdlx5.yerphi.am/ani/ani_collab.html

[32] G. Hovsepyan, L.G. Melkumyan et al., ANI Collaboration report No. 6, YerPhI Preprint 1587/08 (2003).
[33] A.D. Erlykin and E.W. Wolfendale, J. Phys. G, **27** (2001) 941.
[34] D.V. Reames, Space Sci. Rev. **90** (1999) 413.
[35] A.J. Tylka, W.F. Dietrich, C. Lopate and D.V. Reames, Proc. 27th ICRC, Hamburg, (2001).
[36] S. Masuda, T. Kosugi et al., Nature, **371** (1994) 495.
[37] T. Kosugi, K. Makishima et al., Sol. Phys. **136** (1991) 17.
[38] M.J. Aschwanden, M.J. Wills et al., Astrophys. J. **468** (1996) 398.
[39] M.J. Aschwanden, Space Sci. Rev., **101**, No. 1–2, (2002)
[40] Barbosa, Astrophys. J. **425** (1979) 383.
[41] J.A. Miller, N. Guessoum and R. Ramaty, Astrophys. J. **361** (1990) 701.
[42] D.V. Reames, SEPs: Space Weather Hazard in Interplanetary Space, *in* Space Weather, eds. P. Song, H.J. Singer and G.L. Siscoe (Geophysical Monograph 125, AGU,Washington DC, 2002), p. 101.
[43] S. Krucker and R.P. Lin, Astrophys. J. **542** (2000) L61.
[44] R.P. Lin et al., Sol. Phys. **71** (1995) 125.
[45] A.J. Tylka, AGU publication N 2000JA004028, (2001)
[46] A.J. Tylka, C.M.S. Cohen et al., Proc. 27th ICRC, Hamburg, **8** (2001) 3189.
[47] N. Gopalswamy, S. Yashiro et al., Astrophys. J. **572** (2002) L103.
[48] Gaensler et al., (2003), astro-ph/0303427.
[49] A.D. Erlykin and E.W. Wolfendale, J. Phys. G,exit **23** (1997) 979.
[50] A.D. Erlykin and E.W. Wolfendale, Astropart. Phys. **8** (1998) 265.
[51] R. Schlickeiser et al., *Cosmic Ray Astrophysics* (Springer, Berlin, 2002), p. 69.

FLAVOUR PHYSICS AND CP VIOLATION

R. Fleischer
CERN, Geneva, Switzerland

Abstract
The starting point of these lectures is an introduction to the weak interactions of quarks and the Standard-Model description of CP violation, where the key element is the Cabibbo–Kobayashi–Maskawa matrix and the corresponding unitarity triangles. Since the B-meson system will govern the stage of (quark) flavour physics and CP violation in this decade, it will be—after a brief look at the kaon system—our main focus. We shall classify B-meson decays, introduce the theoretical tools to deal with them, explore the requirements for non-vanishing CP-violating asymmetries, and discuss B_q^0–$\bar{B}_q^0$ mixing ($q \in \{d, s\}$). We will then turn to B-factory benchmark modes, discuss the physics potential of B_s^0 mesons, which is particularly promising for B-decay experiments at hadron colliders, and emphasize the importance of studies of rare decays, which are absent at the tree level in the Standard Model, complement nicely the studies of CP violation, and provide interesting probes for new physics.

1 INTRODUCTION

The violation of the CP symmetry, where C and P are the charge-conjugation and parity-transformation operators, respectively, is one of the fundamental and most exciting phenomena in particle physics. Although weak interactions are not invariant under P (and C) transformations, as discovered in 1957, it was believed for several years that the product CP was preserved. Consider, for instance, the process

$$\pi^+ \to e^+ \nu_e \;\xrightarrow{\mathcal{C}}\; \pi^- \to e^- \nu_e^{\rm C} \;\xrightarrow{\mathcal{P}}\; \pi^- \to e^- \bar{\nu}_e, \tag{1}$$

where the left-handed $\nu_e^{\rm C}$ state is not observed in nature; only after performing an additional parity transformation do we obtain the usual right-handed electron antineutrino. Consequently, it appears as if CP was conserved in weak interactions. However, in 1964, it was discovered through the observation of $K_{\rm L} \to \pi^+\pi^-$ decays that weak interactions are *not* invariant under CP transformations [1].

After its discovery, CP violation was, for a very long time, only accessible in the neutral kaon system, where it is described by two complex parameters, ε and ε'; a non-zero value of the latter could only be established—after tremendous efforts—in 1999 [2], [3]. In 2001, CP violation could then also be observed in decays of neutral B mesons [4], [5], which represents the beginning of a new era in the exploration of this phenomenon. Despite this impressive progress, we still have few experimental insights into CP violation, which originates, within the Standard Model (SM) of electroweak interactions, from the flavour structure of the charged-current interactions [6]. One of the main motivations for the exploration of CP violation is that 'new' physics (NP), i.e. physics lying beyond the SM, is typically also associated with new sources of CP violation and new flavour structures [7]– [9]. This is actually the case in many specific NP scenarios, for instance in supersymmetry (SUSY), left–right-symmetric models, and in models with extended Higgs sectors. In this context, it is also interesting to note that the evidence for non-vanishing neutrino masses that we obtained over the last years points towards an origin beyond the SM [10], [11], raising — among other issues — also the question of having CP violation in the neutrino sector, which could be studied, in the more distant future, at dedicated neutrino factories [12].

Interestingly, we may also obtain indirect information on CP violation from cosmology. One of the characteristic features of our Universe is the cosmological baryon asymmetry of $\mathcal{O}(10^{-10})$ [13], [14]. As was pointed out by Sakharov [15], one of the necessary conditions to generate such an asymmetry of the

Universe is—in addition to baryon-number violation and deviations from thermal equilibrium—that the elementary interactions violate CP (and C). Model calculations indicate, however, that the CP violation present in the SM is too small to generate the observed matter–antimatter asymmetry of $\mathcal{O}(10^{-10})$ [16]. It is conceivable that the particular kind of NP underlying the baryon asymmetry is associated with very short-distance scales. In this case, it could not be seen in CP-violating effects in weak meson decays. However, as we have noted above, there are also various scenarios for physics beyond the SM that would affect these processes. Moreover, we do not understand the observed patterns of quark and lepton masses, their mixings and the origin of flavour dynamics in general. It is likely that the NP required to understand these features is also related to new sources of CP violation.

The field of (quark) flavour physics and CP violation is very broad. In this decade, it will be governed by studies of decays of B mesons. The asymmetric e^+e^- B factories operating at the $\Upsilon(4S)$ resonance [17], with their detectors BaBar (SLAC) and Belle (KEK), have already been taking data for a couple of years and have produced plenty of exciting results. Moreover, also hadron colliders have a very promising potential for the exploration of B-meson decays. We may expect first interesting results on several processes from run II of the Tevatron soon [18]. The corresponding channels can then be fully exploited in the era of the LHC, in particular by LHCb (CERN) and BTeV (FNAL) [19]. The great interest in B physics—our main topic—originates from the fact that it provides a very fertile testing ground for the SM picture of flavour physics and CP violation, as we shall see in these lectures. The outline is as follows: in Section 2, we have a closer look at the weak interactions of quarks, discuss the quark-mixing matrix, and introduce the unitarity triangle(s). After giving a brief introduction to the CP violation in the kaon system and making first contact with 'rare' K decays in Section 3, we enter the world of the B mesons in Section 4, where we shall classify their decays, discuss the theoretical tools to deal with them, and investigate the requirements for non-vanishing CP asymmetries. In Section 5, we discuss features of neutral B_q mesons ($q \in \{d, s\}$), including the very important phenomenon of B_q^0–$\bar{B}_q^0$ mixing, and introduce the corresponding CP-violating observables. These considerations then allow us to have a closer look at important benchmark modes for the B factories in Section 6, where we shall also address the current experimental status. In Section 7, we discuss the exploration of CP violation with the help of amplitude relations, whereas we shall focus on the B_s-meson system, which is particularly interesting for B-decay studies at hadron colliders, in Section 8. In Section 9, we emphasize the importance of studies of 'rare' B- and K-meson decays, which are absent at the tree level in the SM, and offer important probes for the search of NP. Finally, we summarize our conclusions in Section 10.

For a collection of detailed textbooks and reviews on CP violation and flavour physics, the reader is referred to Refs. [20]– [26]. Since this field is evolving quickly, I shall also address recent developments that took place after the school in Tsakhkadzor in order to complement the material that I presented there. The data refer to the experimental situation in early 2004.

2 CP VIOLATION IN THE STANDARD MODEL

2.1 Weak interactions of quarks and the quark-mixing matrix

In the framework of the Standard Model of electroweak interactions [6], [27], which is based on the spontaneously broken gauge group

$$SU(2)_{\rm L} \times U(1)_{\rm Y} \stackrel{\rm SSB}{\longrightarrow} U(1)_{\rm em}, \tag{2}$$

CP-violating effects may originate from the charged-current interactions of quarks, having the structure

$$D \to U W^-. \tag{3}$$

Here $D \in \{d, s, b\}$ and $U \in \{u, c, t\}$ denote down- and up-type quark flavours, respectively, whereas the W^- is the usual $SU(2)_{\rm L}$ gauge boson. From a phenomenological point of view, it is convenient to

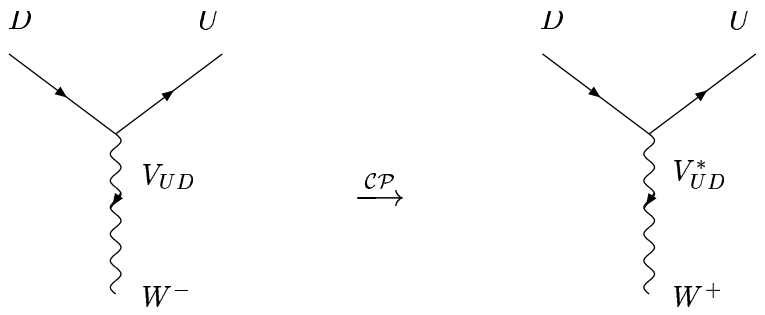


Fig. 1: CP-conjugate charged-current quark-level interaction processes in the SM.

collect the generic 'coupling strengths' V_{UD} of the charged-current processes in (3) in the form of the following matrix:

$$\hat{V}_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{4}$$

which is referred to as the Cabibbo–Kobayashi–Maskawa (CKM) matrix [28], [29].

From a theoretical point of view, this matrix connects the electroweak states (d', s', b') of the down, strange and bottom quarks with their mass eigenstates (d, s, b) through the following unitary transformation [6]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{5}$$

Consequently, $\hat{V}_{\rm CKM}$ is actually a *unitary* matrix. This feature ensures the absence of flavour-changing neutral-current (FCNC) processes at the tree level in the SM, and is hence at the basis of the famous Glashow–Iliopoulos–Maiani (GIM) mechanism [30]. We shall return to the unitarity of the CKM matrix in Subsection 2.6, discussing the 'unitarity triangles'. If we express the non-leptonic charged-current interaction Lagrangian in terms of the mass eigenstates appearing in (5), we arrive at

$$\mathcal{L}^{\rm CC}_{\rm int} = -\frac{g_2}{\sqrt{2}} \left(\begin{array}{cccc} \bar{u}_{\rm L}, & \bar{c}_{\rm L}, & \bar{t}_{\rm L} \end{array} \right) \gamma^\mu \hat{V}_{\rm CKM} \begin{pmatrix} d_{\rm L} \\ s_{\rm L} \\ b_{\rm L} \end{pmatrix} W^\dagger_\mu \; + \; \text{h.c.}, \tag{6}$$

where the gauge coupling g_2 is related to the gauge group $SU(2)_{\rm L}$, and the $W^{(\dagger)}_\mu$ field corresponds to the charged W bosons. Looking at the interaction vertices following from (6), we observe that the elements of the CKM matrix describe in fact the generic strengths of the associated charged-current processes, as we have noted above.

In Fig. 1, we show the $D \to UW^-$ vertex and its CP conjugate. Since the corresponding CP transformation involves the replacement

$$V_{UD} \xrightarrow{\mathcal{CP}} V^*_{UD}, \tag{7}$$

CP violation could—in principle—be accommodated in the SM through complex phases in the CKM matrix. The crucial question in this context is, of course, whether we may actually have physical complex phases in that matrix.

2.2 Phase structure of the CKM matrix

We have the freedom to redefine the up- and down-type quark fields in the following manner:

$$U \to \exp(i\xi_U) U, \quad D \to \exp(i\xi_D) D. \tag{8}$$

If we perform such transformations in (6), the invariance of the charged-current interaction Lagrangian implies the following phase transformations of the CKM matrix elements:

$$V_{UD} \to \exp(i\xi_U) V_{UD} \exp(-i\xi_D). \tag{9}$$

Using these transformations to eliminate unphysical phases, it can be shown that the parametrization of the general $N \times N$ quark-mixing matrix, where N denotes the number of fermion generations, involves the following parameters:

$$\underbrace{\frac{1}{2}N(N-1)}_{\text{Euler angles}} + \underbrace{\frac{1}{2}(N-1)(N-2)}_{\text{complex phases}} = (N-1)^2. \tag{10}$$

If we apply this expression to the case of $N = 2$ generations, we observe that only one rotation angle—the Cabibbo angle $\theta_{\rm C}$ [28]—is required for the parametrization of the 2×2 quark-mixing matrix, which can be written in the following form:

$$\hat{V}_{\rm C} = \begin{pmatrix} \cos\theta_{\rm C} & \sin\theta_{\rm C} \\ -\sin\theta_{\rm C} & \cos\theta_{\rm C} \end{pmatrix}, \tag{11}$$

where $\sin\theta_{\rm C} = 0.22$ can be determined from $K \to \pi\ell\bar{\nu}$ decays. On the other hand, in the case of $N = 3$ generations, the parametrization of the corresponding 3×3 quark-mixing matrix involves three Euler-type angles and a single *complex* phase. This complex phase allows us to accommodate CP violation in the SM, as was pointed out by Kobayashi and Maskawa in 1973 [29]. The corresponding picture is referred to as the Kobayashi–Maskawa (KM) mechanism of CP violation.

In the 'standard parametrization' advocated by the Particle Data Group (PDG) [31], the three-generation CKM matrix takes the following form:

$$\hat{V}_{\rm CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix}, \tag{12}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. Performing appropriate redefinitions of the quark-field phases, the real angles θ_{12}, θ_{23} and θ_{13} can all be made to lie in the first quadrant. The advantage of this parametrization is that the generation labels $i, j = 1, 2, 3$ are introduced in such a manner that the mixing between two chosen generations vanishes if the corresponding mixing angle θ_{ij} is set to zero. In particular, for $\theta_{23} = \theta_{13} = 0$, the third generation decouples, and the 2×2 submatrix describing the mixing between the first and second generations takes the same form as (11).

Another interesting parametrization of the CKM matrix was proposed by Fritzsch and Xing [32]:

$$\hat{V}_{\rm CKM} = \begin{pmatrix} s_{\rm u}s_{\rm d}c + c_{\rm u}c_{\rm d}e^{-i\varphi} & s_{\rm u}c_{\rm d}c - c_{\rm u}s_{\rm d}e^{-i\varphi} & s_{\rm u}s \\ c_{\rm u}s_{\rm d}c - s_{\rm u}c_{\rm d}e^{-i\varphi} & c_{\rm u}c_{\rm d}c + s_{\rm u}s_{\rm d}e^{-i\varphi} & c_{\rm u}s \\ -s_{\rm d}s & -c_{\rm d}s & c \end{pmatrix}. \tag{13}$$

It is inspired by the hierarchical structure of the quark-mass spectrum and is particularly useful in the context of models for fermion masses and mixings. The characteristic feature of this parametrization is that the complex phase arises only in the 2×2 submatrix involving the up, down, strange and charm quarks.

Let us finally note that physical observables, for instance CP-violating asymmetries, *cannot* depend on the chosen parametrization of the CKM matrix, i.e. have to be invariant under the phase transformations specified in (9).

2.3 Further requirements for CP violation

As we have just seen, in order to be able to accommodate CP violation within the framework of the SM through a complex phase in the CKM matrix, at least three generations are required. However, this feature is not sufficient for observable CP-violating effects. To this end, further conditions have to be satisfied, which can be summarized as follows [33], [34]:

$$(m_t^2 - m_c^2)(m_t^2 - m_u^2)(m_c^2 - m_u^2)(m_b^2 - m_s^2)(m_b^2 - m_d^2)(m_s^2 - m_d^2) \times J_{\rm CP} \neq 0, \tag{14}$$

where

$$J_{\rm CP} = |{\rm Im}(V_{i\alpha} V_{j\beta} V_{i\beta}^* V_{j\alpha}^*)| \quad (i \neq j,\, \alpha \neq \beta)\,. \tag{15}$$

The mass factors in (14) are related to the fact that the CP-violating phase of the CKM matrix could be eliminated through an appropriate unitary transformation of the quark fields if any two quarks with the same charge had the same mass. Consequently, the origin of CP violation is closely related to the 'flavour problem' in elementary particle physics, and cannot be understood in a deeper way, unless we have fundamental insights into the hierarchy of quark masses and the number of fermion generations.

The second element of (14), the 'Jarlskog parameter' $J_{\rm CP}$ [33], can be interpreted as a measure of the strength of CP violation in the SM. It does not depend on the chosen quark-field parametrization, i.e. it is invariant under (9), and the unitarity of the CKM matrix implies that all combinations $|{\rm Im}(V_{i\alpha} V_{j\beta} V_{i\beta}^* V_{j\alpha}^*)|$ are equal to one another. Using the standard parametrization of the CKM matrix introduced in (12), we obtain

$$J_{\rm CP} = s_{12} s_{13} s_{23} c_{12} c_{23} c_{13}^2 \sin\delta_{13}. \tag{16}$$

Since the current experimental information on the CKM parameters implies a value of $J_{\rm CP}$ at the 10^{-5} level, CP violation is a small effect in the SM. However, new complex couplings are typically present in scenarios for NP [8], [9], thereby yielding additional sources of CP violation.

2.4 Experimental information on $|V_{\rm CKM}|$

In order to determine the magnitudes $|V_{ij}|$ of the elements of the CKM matrix, we may use the following tree-level processes:

- Nuclear beta decays, neutron decays $\Rightarrow |V_{ud}|$.
- $K \to \pi \ell \bar{\nu}$ decays $\Rightarrow |V_{us}|$.
- ν production of charm off valence d quarks $\Rightarrow |V_{cd}|$.
- Charm-tagged W decays (as well as ν production and semileptonic D decays) $\Rightarrow |V_{cs}|$.
- Exclusive and inclusive $b \to c \ell \bar{\nu}$ decays $\Rightarrow |V_{cb}|$.
- Exclusive and inclusive $b \to u \ell \bar{\nu}$ decays $\Rightarrow |V_{ub}|$.
- $\bar{t} \to \bar{b} \ell \bar{\nu}$ processes $\Rightarrow$ (crude direct determination of) $|V_{tb}|$.

If we use the corresponding experimental information, together with the CKM unitarity condition, and assume that there are only three generations, we arrive at the following 90% C.L. limits for the $|V_{ij}|$ [31]:

$$|\hat{V}_{\rm CKM}| = \begin{pmatrix} 0.9741\text{–}0.9756 & 0.219\text{–}0.226 & 0.0025\text{–}0.0048 \\ 0.219\text{–}0.226 & 0.9732\text{–}0.9748 & 0.038\text{–}0.044 \\ 0.004\text{–}0.014 & 0.037\text{–}0.044 & 0.9990\text{–}0.9993 \end{pmatrix}. \tag{17}$$

In Fig. 2, we have illustrated the resulting hierarchy of the strengths of the charged-current quark-level processes: transitions within the same generation are governed by CKM matrix elements of $\mathcal{O}(1)$, those between the first and the second generation are suppressed by CKM factors of $\mathcal{O}(10^{-1})$, those between the second and the third generation are suppressed by $\mathcal{O}(10^{-2})$, and the transitions between the first and the third generation are even suppressed by CKM factors of $\mathcal{O}(10^{-3})$. In the standard parametrization (12), this hierarchy is reflected by

$$s_{12} = 0.22 \gg s_{23} = \mathcal{O}(10^{-2}) \gg s_{13} = \mathcal{O}(10^{-3}). \tag{18}$$

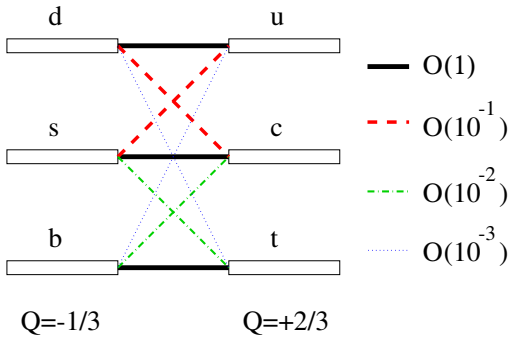


Fig. 2: Hierarchy of the quark transitions mediated through charged-current processes.

2.5 Wolfenstein parametrization of the CKM matrix

For phenomenological applications, it would be useful to have a parametrization of the CKM matrix available that makes the hierarchy arising in (17)—and illustrated in Fig. 2—explicit [35]. In order to derive such a parametrization, we introduce a set of new parameters, λ, A, ρ and η, by imposing the following relations [36]:

$$s_{12} \equiv \lambda = 0.22, \quad s_{23} \equiv A\lambda^2, \quad s_{13}e^{-i\delta_{13}} \equiv A\lambda^3(\rho - i\eta). \tag{19}$$

If we now go back to the standard parametrization (12), we obtain an *exact* parametrization of the CKM matrix as a function of λ (and A, ρ, η), allowing us to expand each CKM element in powers of the small parameter λ. If we neglect terms of $\mathcal{O}(\lambda^4)$, we arrive at the famous 'Wolfenstein parametrization' [35]:

$$\hat{V}_{\rm CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{20}$$

which makes the hierarchical structure of the CKM matrix very transparent and is an important tool for phenomenological considerations, as we shall see throughout these lectures.

For several applications, next-to-leading order corrections in λ play an important role. Using the exact parametrization following from (12) and (19), they can be calculated straightforwardly by expanding each CKM element to the desired accuracy in λ [36], [37]:

$$\begin{aligned} V_{ud} &= 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 + \mathcal{O}(\lambda^6), \quad V_{us} = \lambda + \mathcal{O}(\lambda^7), \quad V_{ub} = A\lambda^3(\rho - i\,\eta), \\ V_{cd} &= -\lambda + \frac{1}{2}A^2\lambda^5\left[1 - 2(\rho + i\eta)\right] + \mathcal{O}(\lambda^7), \\ V_{cs} &= 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4(1 + 4A^2) + \mathcal{O}(\lambda^6), \\ V_{cb} &= A\lambda^2 + \mathcal{O}(\lambda^8), \quad V_{td} = A\lambda^3\left[1 - (\rho + i\eta)\left(1 - \frac{1}{2}\lambda^2\right)\right] + \mathcal{O}(\lambda^7), \\ V_{ts} &= -A\lambda^2 + \frac{1}{2}A(1 - 2\rho)\lambda^4 - i\eta A\lambda^4 + \mathcal{O}(\lambda^6), \quad V_{tb} = 1 - \frac{1}{2}A^2\lambda^4 + \mathcal{O}(\lambda^6). \end{aligned} \tag{21}$$

It should be noted that

$$V_{ub} \equiv A\lambda^3(\rho - i\eta) \tag{22}$$

receives *by definition* no power corrections in λ within this prescription. If we follow [36] and introduce the generalized Wolfenstein parameters

$$\bar{\rho} \equiv \rho\left(1 - \frac{1}{2}\lambda^2\right), \quad \bar{\eta} \equiv \eta\left(1 - \frac{1}{2}\lambda^2\right), \tag{23}$$

we may simply write, up to corrections of $\mathcal{O}(\lambda^7)$,

$$V_{td} = A\lambda^3(1 - \bar{\rho} - i\,\bar{\eta}). \tag{24}$$

Moreover, we have to an excellent accuracy

$$V_{us} = \lambda \quad \text{and} \quad V_{cb} = A\lambda^2, \tag{25}$$

as these quantities receive only corrections at the λ^7 and λ^8 levels, respectively. In comparison with other generalizations of the Wolfenstein parametrization found in the literature, the advantage of (21) is the absence of relevant corrections to V_{us} and V_{cb}, and that V_{ub} and V_{td} take forms similar to those in (20). As far as the Jarlskog parameter introduced in (15) is concerned, we obtain the simple expression

$$J_{\rm CP} = \lambda^6 A^2 \eta, \tag{26}$$

which should be compared with (16).

2.6 Unitarity triangles of the CKM matrix

The unitarity of the CKM matrix, which is described by

$$\hat{V}_{\rm CKM}^{\dagger} \cdot \hat{V}_{\rm CKM} = \hat{1} = \hat{V}_{\rm CKM} \cdot \hat{V}_{\rm CKM}^{\dagger}, \tag{27}$$

leads to a set of 12 equations, consisting of 6 normalization and 6 orthogonality relations. The latter can be represented as 6 triangles in the complex plane [38], all having the same area, $2A_\Delta = J_{\rm CP}$ [39]. Let us now have a closer look at these relations: those describing the orthogonality of different columns of the CKM matrix are given by

$$\underbrace{V_{ud}V_{us}^*}_{\mathcal{O}(\lambda)} + \underbrace{V_{cd}V_{cs}^*}_{\mathcal{O}(\lambda)} + \underbrace{V_{td}V_{ts}^*}_{\mathcal{O}(\lambda^5)} = 0 \tag{28}$$

$$\underbrace{V_{us}V_{ub}^*}_{\mathcal{O}(\lambda^4)} + \underbrace{V_{cs}V_{cb}^*}_{\mathcal{O}(\lambda^2)} + \underbrace{V_{ts}V_{tb}^*}_{\mathcal{O}(\lambda^2)} = 0 \tag{29}$$

$$\underbrace{V_{ud}V_{ub}^*}_{(\rho+i\eta)A\lambda^3} + \underbrace{V_{cd}V_{cb}^*}_{-A\lambda^3} + \underbrace{V_{td}V_{tb}^*}_{(1-\rho-i\eta)A\lambda^3} = 0, \tag{30}$$

whereas those associated with the orthogonality of different rows take the following form:

$$\underbrace{V_{ud}^*V_{cd}}_{\mathcal{O}(\lambda)} + \underbrace{V_{us}^*V_{cs}}_{\mathcal{O}(\lambda)} + \underbrace{V_{ub}^*V_{cb}}_{\mathcal{O}(\lambda^5)} = 0 \tag{31}$$

$$\underbrace{V_{cd}^*V_{td}}_{\mathcal{O}(\lambda^4)} + \underbrace{V_{cs}^*V_{ts}}_{\mathcal{O}(\lambda^2)} + \underbrace{V_{cb}^*V_{tb}}_{\mathcal{O}(\lambda^2)} = 0 \tag{32}$$

$$\underbrace{V_{ud}^*V_{td}}_{(1-\rho-i\eta)A\lambda^3} + \underbrace{V_{us}^*V_{ts}}_{-A\lambda^3} + \underbrace{V_{ub}^*V_{tb}}_{(\rho+i\eta)A\lambda^3} = 0. \tag{33}$$

Here we have also indicated the structures that arise if we apply the Wolfenstein parametrization by keeping just the leading, non-vanishing terms. We observe that only in (30) and (33), which describe the orthogonality of the first and third columns and of the first and third rows, respectively, are all three sides of comparable magnitude, $\mathcal{O}(\lambda^3)$, while in the remaining relations, one side is suppressed with respect to the others by factors of $\mathcal{O}(\lambda^2)$ or $\mathcal{O}(\lambda^4)$. Consequently, we have to deal with only *two* non-squashed unitarity triangles in the complex plane. However, as we have already indicated in (30) and (33), the corresponding orthogonality relations agree with each other at the λ^3 level, yielding

$$[(\rho + i\eta) + (-1) + (1 - \rho - i\eta)]\, A\lambda^3 = 0. \tag{34}$$

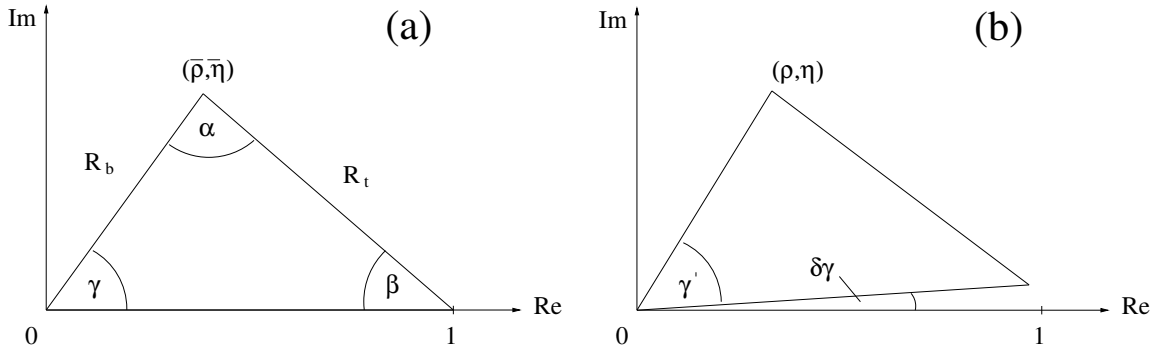


Fig. 3: The two non-squashed unitarity triangles of the CKM matrix, as explained in the text: (a) and (b) correspond to the orthogonality relations (30) and (33), respectively.

Consequently, they describe the same triangle, which is usually referred to as *the* unitarity triangle of the CKM matrix [39], [40].

In the second-generation B-decay experiments at the LHC, the experimental accuracy will be so tremendous that we shall also have to take the next-to-leading order terms of the Wolfenstein expansion into account, and shall have to distinguish between the unitarity triangles following from (30) and (33). Let us first have a closer look at the former relation. Including terms of $\mathcal{O}(\lambda^5)$, we obtain the following generalization of (34):

$$\left[(\bar{\rho}+i\bar{\eta})+(-1)+(1-\bar{\rho}-i\bar{\eta})\right]A\lambda^3+\mathcal{O}(\lambda^7)=0, \tag{35}$$

where $\bar{\rho}$ and $\bar{\eta}$ are as defined in (23). If we divide this relation by the overall normalization factor $A\lambda^3$, and introduce

$$R_b\equiv\sqrt{\overline{\rho}^2+\overline{\eta}^2}=\left(1-\frac{\lambda^2}{2}\right)\frac{1}{\lambda}\left|\frac{V_{ub}}{V_{cb}}\right| \tag{36}$$

$$R_t\equiv\sqrt{(1-\overline{\rho})^2+\overline{\eta}^2}=\frac{1}{\lambda}\left|\frac{V_{td}}{V_{cb}}\right|, \tag{37}$$

we arrive at the unitarity triangle illustrated in Fig. 3 (a). It is a straightforward generalization of the leading-order case described by (34): instead of (ρ,η), the apex is now simply given by $(\bar{\rho},\bar{\eta})$ [36]. The two sides R_b and R_t, as well as the three angles α, β and γ, will show up at several places throughout these lectures. Moreover, the relations

$$V_{ub}=A\lambda^3\left(\frac{R_b}{1-\lambda^2/2}\right)e^{-i\gamma},\quad V_{td}=A\lambda^3R_te^{-i\beta} \tag{38}$$

are also useful for phenomenological applications, since they make the dependences of γ and β explicit; they correspond to the phase convention chosen both in the standard parametrization (12) and in the generalized Wolfenstein parametrization (21). Finally, if we take also (19) into account, we obtain

$$\delta_{13}=\gamma. \tag{39}$$

Let us now turn to (33). Here we arrive at an expression that is more complicated than (35):

$$\left[\left\{1-\frac{\lambda^2}{2}-(1-\lambda^2)\rho-i(1-\lambda^2)\eta\right\}+\left\{-1+\left(\frac{1}{2}-\rho\right)\lambda^2-i\eta\lambda^2\right\}+\{\rho+i\eta\}\right]A\lambda^3+\mathcal{O}(\lambda^7)=0. \tag{40}$$

If we divide again by $A\lambda^3$, we obtain the unitarity triangle sketched in Fig. 3 (b), where the apex is given by (ρ,η) and *not* by $(\bar{\rho},\bar{\eta})$. On the other hand, we encounter a tiny angle

$$\delta\gamma\equiv\lambda^2\eta=\mathcal{O}(1^\circ) \tag{41}$$

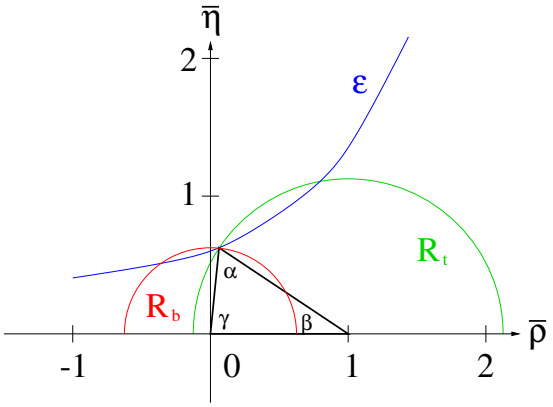


Fig. 4: Contours in the $\bar\rho$–$\bar\eta$ plane, allowing us to determine the apex of the UT.

between real axis and basis of the triangle, which satisfies

$$\gamma = \gamma' + \delta\gamma, \tag{42}$$

where γ coincides with the corresponding angle in Fig. 3 (a).

Whenever we refer to a 'unitarity triangle' (UT) in the following discussion, we mean the one illustrated in Fig. 3 (a), which is the generic generalization of the leading-order case described by (34). As we shall see below, the UT is the central target of the experimental tests of the SM description of CP violation. Interestingly, also the tiny angle $\delta\gamma$ can be probed directly through certain CP-violating effects that can be explored at hadron colliders, in particular at the LHC.

2.7 Towards an allowed region in the $\bar\rho$–$\bar\eta$ plane

It is possible to constrain—and even determine—the apex of the UT in the $\bar\rho$–$\bar\eta$ plane with the help of experimental data. Unfortunately, we do not yet have the theoretical framework available to discuss in detail how this can actually be done (but this will become obvious in the course of these lectures). However, it is nevertheless useful to sketch the corresponding procedure—the 'CKM fits'—already now, consisting of the following elements:

- The parameter R_b introduced in (36), which involves the ratio $|V_{ub}/V_{cb}|$. It can be determined experimentally through $b \to u\ell\bar\nu$ and $b \to c\ell\bar\nu$ decay processes. Following these lines, we may fix a circle in the $\bar\rho$–$\bar\eta$ plane that is centred at the origin $(0,0)$ and has the radius R_b.
- The parameter R_t introduced in (37), which involves the ratio $|V_{td}/V_{cb}|$. It can be determined with the help of the mass differences $\Delta M_{d,s}$ of the mass eigenstates of the neutral B_d- and B_s-meson systems. Experimental information on these quantities then allows us to fix another circle in the $\bar\rho$–$\bar\eta$ plane, which is centred at $(1,0)$ and has the radius R_t.
- Finally, we may convert the measurement of the observable ε, which describes the CP violation in the neutral kaon system that was discovered in 1964, into a hyperbola in the $\bar\rho$–$\bar\eta$ plane.

In Fig. 4, we have illustrated these contours; their intersection allows us to determine the apex of the UT within the SM. The curves that are implied by ΔM_d and ε depend on the CKM parameter A and the top-quark mass m_t, as well as on certain perturbatively calculable QCD corrections and non-perturbative parameters. Consequently, strong correlations between the theoretical and experimental uncertainties arise in the CKM fits. As discussed in detail in Ref. [41], several different approaches can be found in the literature to deal with the corresponding error propagation. The typical (conservative) ranges for the UT angles that follow from the CKM fits read as follows:

$$70^\circ \lesssim \alpha \lesssim 130^\circ, \quad 20^\circ \lesssim \beta \lesssim 30^\circ, \quad 50^\circ \lesssim \gamma \lesssim 70^\circ. \tag{43}$$

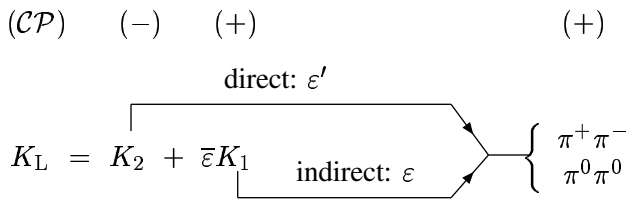


Fig. 5: Illustration of indirect and direct CP violation in $K_{\rm L} \to \pi\pi$ decays.

On the other hand, CP violation in the B-meson system provides various strategies to determine these angles *directly*, thereby offering different ways to fix the apex of the UT in the $\bar{\rho}$–$\bar{\eta}$ plane. Following these lines, a powerful test of the KM mechanism can be performed. This very interesting feature is also reflected by the tremendous efforts to explore CP violation in B decays experimentally in this decade. Before having a closer look at B mesons, their decays, the theoretical tools to deal with them and the general requirements for having non-vanishing CP asymmetries, let us first turn to the kaon system.

3 A FIRST LOOK AT CP VIOLATION AND RARE DECAYS IN THE KAON SYSTEM

3.1 CP violation: ε and ε'

As we have already noted, in 1964, CP violation was discovered—as a big surprise—in the famous experiment by Christenson et al. [1], who observed $K_{\rm L} \to \pi^+\pi^-$ decays. If the weak interactions *were* invariant under CP transformations, the mass eigenstates $K_{\rm S}$ and $K_{\rm L}$ of the Hamilton operator describing K^0–$\bar{K}^0$ mixing *were* eigenstates of the CP operator, with eigenvalues $+1$ and -1, respectively. Since the $\pi^+\pi^-$ final state of $K_{\rm L} \to \pi^+\pi^-$ is CP-even, the detection of this transition signals indeed the violation of the CP symmetry in weak interaction processes. The discussion in this subsection serves mainly to make a first contact with this phenomenon; for detailed presentations of CP violation in kaon decays, we refer the reader to Refs. [21], [22], [37].

In the neutral K-meson system, CP violation is described by two complex quantities, called ε and ε', which are defined by the following ratios of decay amplitudes:

$$\frac{A(K_{\rm L} \to \pi^+\pi^-)}{A(K_{\rm S} \to \pi^+\pi^-)} \approx \varepsilon + \varepsilon', \quad \frac{A(K_{\rm L} \to \pi^0\pi^0)}{A(K_{\rm S} \to \pi^0\pi^0)} \approx \varepsilon - 2\,\varepsilon'. \tag{44}$$

These parameters are associated with 'indirect' and 'direct' CP violation, as we have illustrated in Fig. 5, where K_1 and K_2 denote the CP eigenstates of the neutral kaon system with CP eigenvalues $+1$ and -1, respectively. The terminology of 'indirect CP violation' originates from the fact that the mass eigenstate $K_{\rm L}$ of the neutral kaon system is *not* an eigenstate of the CP operator because of the small admixture of the CP-even K_1 state, which may decay—through a CP-conserving transition—into a $\pi\pi$ final state. On the other hand, direct CP violation originates from *direct* transitions of the CP-odd K_2 state into the CP-even $\pi\pi$ final state.

After the discovery of indirect CP violation through $K_{\rm L} \to \pi^+\pi^-$ decays, this phenomenon could also be observed in $K_{\rm L} \to \pi^0\pi^0$, $\pi\ell\bar{\nu}$, $\pi^+\pi^-\gamma$ modes, and recently in $K_{\rm L} \to \pi^+\pi^- e^+e^-$ transitions. All these effects can be described by

$$\varepsilon = (2.280 \pm 0.013) \times e^{i\frac{\pi}{4}} \times 10^{-3}. \tag{45}$$

As we noted in Subsection 2.7, the knowledge of the CKM parameter A and the top-quark mass m_t allows us—in combination with the calculation of perturbative QCD corrections and estimates of non-perturbative parameters—to convert the observable ε into a hyperbola in the $\bar{\rho}$–$\bar{\eta}$ plane, as is explicitly shown in Refs. [21], [22], [37]. This analysis implies in particular $\bar{\eta} > 0$, i.e. that the apex of the UT lies in the *upper* half of the $\bar{\rho}$–$\bar{\eta}$ plane.

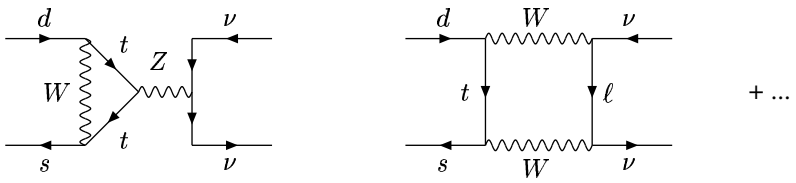


Fig. 6: Decay processes contributing to $K_{\rm L} \to \pi^0 \nu\bar{\nu}$ in the SM.

Direct CP violation in neutral $K \to \pi\pi$ decays can be described through the quantity $\mathrm{Re}(\varepsilon'/\varepsilon)$. In 1999, measurements at CERN (NA48) [2] and FNAL (KTeV) [3] have demonstrated—after tremendous efforts over many years—that this observable is actually *different* from zero, thereby establishing the phenomenon of *direct* CP violation. The experimental status is now given as follows:

$$\mathrm{Re}(\varepsilon'/\varepsilon) = \begin{cases} (14.7 \pm 2.2) \times 10^{-4} & \text{(NA48 [42])}, \\ (20.7 \pm 2.8) \times 10^{-4} & \text{(KTeV [43])}. \end{cases} \tag{46}$$

If we take also the previous results of the NA31 and E731 collaborations into account, we obtain the world average

$$\mathrm{Re}(\varepsilon'/\varepsilon) = (16.6 \pm 1.6) \times 10^{-4}. \tag{47}$$

Within the SM, calculations of $\mathrm{Re}(\varepsilon'/\varepsilon)$ give the same order of magnitude (for an overview of the current status, see Ref. [44]). However, these analyses are affected by large hadronic uncertainties; the situation is particularly unfavourable, since $\mathrm{Re}(\varepsilon'/\varepsilon)$ is governed by the competition between two different decay topologies and suffers from a strong cancellation between them. Consequently, although the measurement of $\mathrm{Re}(\varepsilon'/\varepsilon)$ led to the discovery of a new kind of CP violation, this observable does unfortunately not allow us to perform stringent tests of the KM mechanism of CP violation, unless better techniques to deal with the hadronic uncertainties are available.

3.2 Rare decays: $K \to \pi\nu\bar{\nu}$

From a theoretical point of view, the decays $K_{\rm L} \to \pi^0 \nu\bar{\nu}$ and $K^+ \to \pi^+ \nu\bar{\nu}$ are very interesting. Since we shall have a detailed look at them in Subsection 9.3, let us here just sketch their most interesting features. As can easily be seen, these transitions originate from FCNC processes. Consequently, because of the GIM mechanism, they receive no contributions at the tree level in the SM. However, they may be induced through loop processes of the kind shown in Fig. 6, and are therefore strongly suppressed transitions, which are referred to as 'rare' decays. One of the most exciting features of the $K \to \pi\nu\bar{\nu}$ modes is that they are theoretically very clean. Moreover, it can be shown that the measurement of the $K_{\rm L} \to \pi^0 \nu\bar{\nu}$ branching ratio allows us to determine $|\bar{\eta}|$, whereas the one of $K^+ \to \pi^+ \nu\bar{\nu}$ can be converted into an ellipse in the $\bar{\rho}$–$\bar{\eta}$ plane. The intersection of these contours provides an interesting determination of the UT, where in particular $\sin 2\beta$ can be extracted with respectable accuracy [45]. We may hence perform a stringent test of the SM description of CP violation by comparing the UT thus determined with the ones following from the construction illustrated in Fig. 4 and the studies of CP violation in the B-meson system. In particular, as we shall see in Subsection 6.1, $B_d \to J/\psi K_{\rm S}$ decays allow also a clean determination of $\sin 2\beta$, so that a violation of the SM relation

$$(\sin 2\beta)_{\pi\nu\bar{\nu}} = (\sin 2\beta)_{\psi K_{\rm S}} \tag{48}$$

would indicate sources of CP violation lying beyond the SM. Moreover, also the determination of the angle γ of the UT is interesting for the search of NP with $K \to \pi\nu\bar{\nu}$ decays [46], [47].

Unfortunately, the $K \to \pi\nu\bar{\nu}$ branching ratios are extremely small. A recent update of the corresponding calculations within the SM yields the following results [48]:

$$\mathrm{BR}(K^+ \to \pi^+ \nu\bar{\nu}) = (8.0 \pm 1.1) \times 10^{-11}, \quad \mathrm{BR}(K_{\rm L} \to \pi^0 \nu\bar{\nu}) = (3.2 \pm 0.6) \times 10^{-11}, \tag{49}$$

which are in the ballpark of other recent analyses [49], [50]. Interestingly, a third event for the former channel was very recently observed by the E949 experiment at BNL [51], thereby complementing the previous observation of the two events by the E787 collaboration [52]. The three observed $K^+ \to \pi^+\nu\bar{\nu}$ events can be converted into the following branching ratio:

$$\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu}) = (14.7^{+13.0}_{-8.9}) \times 10^{-11}. \tag{50}$$

On the other hand, for the $K_{\rm L} \to \pi^0\nu\bar{\nu}$ channel, only the experimental upper bound

$$\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu}) < 5.9 \times 10^{-7} \tag{51}$$

is available from the KTeV collaboration [53].

In the presence of NP, the $K \to \pi\nu\bar{\nu}$ branching ratios may differ strongly from the SM expectations given in (49). For instance, in a recent NP analysis [48], [54], which is motivated by certain puzzling patterns in the B-factory data and will be discussed in Subsection 9.4, a spectacular enhancement of the $K_{\rm L} \to \pi^0\nu\bar{\nu}$ branching ratio, by one order of magnitude, is found, and the relation in (48) would in fact be dramatically violated.

Concerning the experimental aspects of the $K \to \pi\nu\bar{\nu}$ modes, we refer the reader to the recent overview given in Ref. [55]. Let us now move on to the central topic of these lectures, the B-meson system.

4 DECAYS OF B MESONS

The B-meson system consists of charged and neutral B mesons, which are characterized by the

$$\begin{array}{ll} B^+ \sim u\,\bar{b}, & B^- \sim \bar{u}\,b \\ B_c^+ \sim c\,\bar{b}, & B_c^+ \sim \bar{c}\,b \end{array}$$

and

$$\begin{array}{ll} B_d^0 \sim d\,\bar{b}, & \bar{B}_d^0 \sim \bar{d}\,b \\ B_s^0 \sim s\,\bar{b}, & \bar{B}_s^0 \sim \bar{s}\,b \end{array}$$

valence-quark contents, respectively. The characteristic feature of the neutral B_q ($q \in \{d, s\}$) mesons is the phenomenon of B_q^0–$\bar{B}_q^0$ mixing (the counterpart of K^0–$\bar{K}^0$ mixing), which will be discussed in Subsection 5.1. As far as the weak decays of B mesons are concerned, we distinguish between leptonic, semileptonic and non-leptonic transitions.

4.1 Leptonic decays

The simplest B-meson decay class is given by leptonic decays of the kind $B^- \to \ell\bar{\nu}$, as illustrated in Fig. 7. If we evaluate the corresponding Feynman diagram, we arrive at the following transition amplitude:

$$T_{fi} = -\frac{g_2^2}{8} V_{ub} \underbrace{[\bar{u}_\ell \gamma^\alpha (1-\gamma_5) v_\nu]}_{\text{Dirac spinors}} \left[\frac{g_{\alpha\beta}}{k^2 - M_W^2}\right] \underbrace{\langle 0|\bar{u}\gamma^\beta(1-\gamma_5)b|B^-\rangle}_{\text{hadronic ME}}, \tag{52}$$

where g_2 is the $SU(2)_{\rm L}$ gauge coupling, V_{ub} the corresponding element of the CKM matrix, α and β are Lorentz indices, and M_W denotes the mass of the W gauge boson. Since the four-momentum k that is carried by the W satisfies $k^2 = M_B^2 \ll M_W^2$, we may write

$$\frac{g_{\alpha\beta}}{k^2 - M_W^2} \quad \longrightarrow \quad -\frac{g_{\alpha\beta}}{M_W^2} \equiv -\left(\frac{8G_{\rm F}}{\sqrt{2}g_2^2}\right) g_{\alpha\beta}, \tag{53}$$

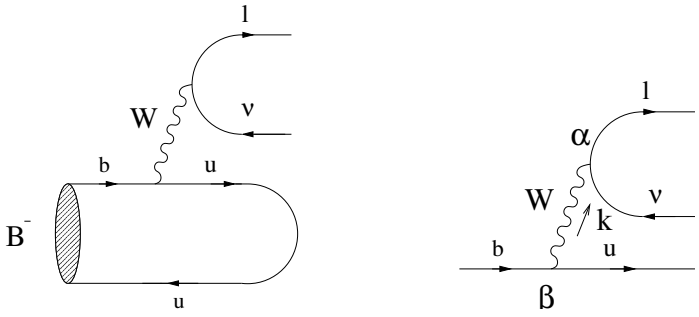


Fig. 7: Feynman diagram contributing to the leptonic decay $B^- \to \ell\bar{\nu}$.

where $G_{\rm F}$ is Fermi's constant. Consequently, we may 'integrate out' the W boson in (52), which yields

$$T_{fi} = \frac{G_{\rm F}}{\sqrt{2}} V_{ub} \left[\bar{u}_\ell \gamma^\alpha (1-\gamma_5) v_\nu\right] \langle 0|\bar{u}\gamma_\alpha(1-\gamma_5)b|B^-\rangle. \tag{54}$$

In this simple expression, *all* the hadronic physics is encoded in the *hadronic matrix element*

$$\langle 0|\bar{u}\gamma_\alpha(1-\gamma_5)b|B^-\rangle,$$

i.e. there are no other strong-interaction (QCD) effects. Since the B^- meson is a pseudoscalar particle, we have

$$\langle 0|\overline{u}\gamma_\alpha b|B^-\rangle = 0, \tag{55}$$

and may write

$$\langle 0|\bar{u}\gamma_\alpha\gamma_5 b|B^-(q)\rangle = i f_B q_\alpha, \tag{56}$$

where f_B is the B-meson *decay constant*, which is an important input for phenomenological studies. In order to determine this quantity, which is a very challenging task, non-perturbative techniques, such as lattice [56] or QCD sum-rule analyses [57], are required. If we use (54) with (55) and (56), and perform the corresponding phase-space integrations, we obtain the following decay rate:

$$\Gamma(B^- \to \ell\bar{\nu}) = \frac{G_{\rm F}^2}{8\pi} |V_{ub}|^2 M_B m_\ell^2 \left(1 - \frac{m_\ell^2}{M_B^2}\right)^2 f_B^2, \tag{57}$$

where M_B and m_ℓ denote the masses of the B^- and ℓ, respectively. Because of the tiny value of $|V_{ub}| \propto \lambda^3$ and a helicity-suppression mechanism, we obtain unfortunately very small branching ratios of $\mathcal{O}(10^{-10})$ and $\mathcal{O}(10^{-7})$ for $\ell = e$ and $\ell = \mu$, respectively [58]. The helicity suppression is not effective for $\ell = \tau$, but—because of the required τ reconstruction—these modes are also very challenging from an experimental point of view. A measurement of leptonic B-meson decays would nevertheless be very interesting, as it would allow an experimental determination of f_B, thereby providing tests of non-perturbative calculations of this important parameter.[1] The CKM element $|V_{ub}|$ can be extracted from semileptonic B decays, our next topic.

4.2 Semileptonic decays

4.2.1 General structure

Semileptonic B-meson decays of the kind shown in Fig. 8 have a structure that is more complicated than the one of the leptonic transitions. If we evaluate the corresponding Feynman diagram for the $b \to c$

[1] Leptonic decays of $D_{(s)}$ mesons allow the extraction of the corresponding decay constants $f_{D_{(s)}}$, which are defined in analogy to (56). These measurements are an important element of the CLEO-c research programme [59].

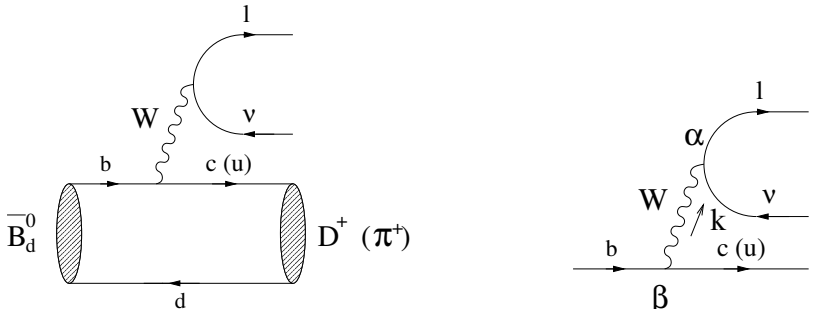


Fig. 8: Feynman diagram contributing to semileptonic $\bar{B}_d^0 \to D^+(\pi^+)\ell\bar{\nu}$ decays.

case, we obtain

$$T_{fi} = -\frac{g_2^2}{8} V_{cb} \underbrace{[\bar{u}_\ell \gamma^\alpha (1-\gamma_5) v_\nu]}_{\text{Dirac spinors}} \left[\frac{g_{\alpha\beta}}{k^2 - M_W^2}\right] \underbrace{\langle D^+ | \bar{c}\gamma^\beta (1-\gamma_5) b | \bar{B}_d^0 \rangle}_{\text{hadronic ME}}. \tag{58}$$

Because of $k^2 \sim M_B^2 \ll M_W^2$, we may again—as in (52)—integrate out the W boson with the help of (53), which yields

$$T_{fi} = \frac{G_{\rm F}}{\sqrt{2}} V_{cb} \, [\bar{u}_\ell \gamma^\alpha (1-\gamma_5) v_\nu] \, \langle D^+ | \bar{c}\gamma_\alpha (1-\gamma_5) b | \bar{B}_d^0 \rangle, \tag{59}$$

where *all* the hadronic physics is encoded in the hadronic matrix element

$$\langle D^+ | \bar{c}\gamma_\alpha (1-\gamma_5) b | \bar{B}_d^0 \rangle,$$

i.e. there are *no* other strong-interaction (QCD) effects. Since the $\bar{B}_d^0$ and D^+ are pseudoscalar mesons, we have

$$\langle D^+ | \bar{c}\gamma_\alpha \gamma_5 b | \bar{B}_d^0 \rangle = 0, \tag{60}$$

and may write

$$\langle D^+(k) | \bar{c}\gamma_\alpha b | \bar{B}_d^0(p) \rangle = F_1(q^2) \left[(p+k)_\alpha - \left(\frac{M_B^2 - M_D^2}{q^2}\right) q_\alpha\right] + F_0(q^2) \left(\frac{M_B^2 - M_D^2}{q^2}\right) q_\alpha, \tag{61}$$

where $q \equiv p - k$, and the $F_{1,0}(q^2)$ denote the *form factors* of the $\bar{B} \to D$ transitions. Consequently, in contrast to the simple case of the leptonic transitions, semileptonic decays involve *two* hadronic form factors instead of the decay constant f_B. In order to calculate these parameters, which depend on the momentum transfer q, again non-perturbative techniques (lattice, QCD sum rules, etc.) are required.

4.2.2 Aspects of the heavy-quark effective theory

If the mass m_Q of a quark Q is much larger than the QCD scale parameter $\Lambda_{\rm QCD} = \mathcal{O}(100\,\text{MeV})$, it is referred to as a 'heavy' quark. Since the bottom and charm quarks have masses at the level of 5 GeV and 1 GeV, respectively, they belong to this important category. As far as the extremely heavy top quark, with $m_t \sim 170$ GeV is concerned, it decays unfortunately through weak interactions before a hadron can be formed. Let us now consider a heavy quark that is bound inside a hadron, i.e. a bottom or a charm quark. The heavy quark then moves almost with the hadron's four velocity v and is almost on-shell, so that

$$p_Q^\mu = m_Q v^\mu + k^\mu, \tag{62}$$

where $v^2 = 1$ and $k \ll m_Q$ is the 'residual' momentum. Owing to the interactions of the heavy quark with the light degrees of freedom of the hadron, the residual momentum may only change by $\Delta k \sim \Lambda_{\rm QCD}$, and $\Delta v \to 0$ for $\Lambda_{\rm QCD}/m_Q \to 0$.

It is now instructive to have a look at the elastic scattering process $\bar{B}(v) \to \bar{B}(v')$ in the limit of $\Lambda_{\rm QCD}/m_b \to 0$, which is characterized by the following matrix element:

$$\frac{1}{M_B}\langle \bar{B}(v')|\bar{b}_{v'}\gamma_\alpha b_v|\bar{B}(v)\rangle = \xi(v'\cdot v)(v+v')_\alpha. \tag{63}$$

Since the contraction of this matrix element with $(v-v')^\alpha$ has to vanish because of $\not{v} b_v = b_v$ and $\overline{b}_{v'}\not{v}' = \overline{b}_{v'}$, no $(v-v')_\alpha$ term arises in the parametrization in (63). On the other hand, the $1/M_B$ factor is related to the normalization of states, i.e. the right-hand side of

$$\left(\frac{1}{\sqrt{M_B}}\langle \bar{B}(p')|\right)\left(|\bar{B}(p)\rangle\frac{1}{\sqrt{M_B}}\right) = 2v^0(2\pi)^3\delta^3(\vec{p}-\vec{p}) \tag{64}$$

does not depend on M_B. Finally, current conservation implies the following normalization condition:

$$\xi(v'\cdot v = 1) = 1, \tag{65}$$

where the 'Isgur–Wise' function $\xi(v'\cdot v)$ does *not* depend on the flavour of the heavy quark (heavy-quark symmetry) [60]. Consequently, for $\Lambda_{\rm QCD}/m_{b,c} \to 0$, we may write

$$\frac{1}{\sqrt{M_D M_B}}\langle D(v')|\bar{c}_{v'}\gamma_\alpha b_v|\bar{B}(v)\rangle = \xi(v'\cdot v)(v+v')_\alpha, \tag{66}$$

and observe that this transition amplitude is governed—in the heavy-quark limit—by *one* hadronic form factor $\xi(v'\cdot v)$, which satisfies $\xi(1) = 1$. If we now compare (66) with (61), we obtain

$$F_1(q^2) = \frac{M_D + M_B}{2\sqrt{M_D M_B}}\xi(w) \tag{67}$$

$$F_0(q^2) = \frac{2\sqrt{M_D M_B}}{M_D + M_B}\left[\frac{1+w}{2}\right]\xi(w), \tag{68}$$

with

$$w \equiv v_D \cdot v_B = \frac{M_D^2 + M_B^2 - q^2}{2M_D M_B}. \tag{69}$$

Similar relations hold also for the $\bar{B} \to D^*$ form factors because of the heavy-quark spin symmetry, since the D^* is related to the D by a rotation of the heavy-quark spin. A detailed discussion of these interesting features and the associated 'heavy-quark effective theory' (HQET) is beyond the scope of these lectures. For a detailed overview, we refer the reader to Ref. [61], where also a comprehensive list of the original references can be found. For a more phenomenological discussion, also Ref. [17] is very useful.

4.2.3 Applications

An important application of the formalism sketched above is the extraction of the CKM element $|V_{cb}|$. To this end, $\bar{B} \to D^*\ell\bar{\nu}$ decays are particularly promising. The corresponding rate can be written as

$$\frac{{\rm d}\Gamma}{{\rm d}w} = G_{\rm F}^2 K(M_B, M_{D^*}, w) F(w)^2 |V_{cb}|^2, \tag{70}$$

where $K(M_B, M_{D^*}, w)$ is a known kinematic function, and $F(w)$ agrees with the Isgur–Wise function, up to perturbative QCD corrections and $\Lambda_{\rm QCD}/m_{b,c}$ terms. The form factor $F(w)$ is a non-perturbative quantity. However, it satisfies the following normalization condition:

$$F(1) = \eta_A(\alpha_s)\left[1 + \frac{0}{m_c} + \frac{0}{m_b} + \mathcal{O}(\Lambda_{\rm QCD}^2/m_{b,c}^2)\right], \tag{71}$$

where $\eta_A(\alpha_s)$ is a perturbatively calculable short-distance QCD factor, and the $\Lambda_{\rm QCD}/m_{b,c}$ corrections *vanish* [61, 62]. The important latter feature is an implication of Luke's theorem [63]. Consequently, if we extract $F(w)|V_{cb}|$ from a measurement of (70) as a function of w and extrapolate to the 'zero-recoil point' $w = 1$ (where the rate vanishes), we may determine $|V_{cb}|$. In the case of $\bar{B} \to D\ell\bar{\nu}$ decays, we have $\mathcal{O}(\Lambda_{\rm QCD}/m_{b,c})$ corrections to the corresponding rate $\mathrm{d}\Gamma/\mathrm{d}w$ at $w = 1$. In order to determine $|V_{cb}|$, inclusive $B \to X_c\ell\bar{\nu}$ decays also offer very attractive avenues. As becomes obvious from (25) and the considerations in Subsection 2.6, $|V_{cb}|$ fixes the normalization of the UT. Moreover, this quantity is an important input parameter for various theoretical calculations. Its current experimental status can be summarized as follows:

$$|V_{cb}| = 0.04 \times [1 \pm 0.05] \quad \Rightarrow \quad A = 0.83 \times [1 \pm 0.05]. \tag{72}$$

Let us now turn to $\bar{B} \to \pi\ell\bar{\nu}, \rho\ell\bar{\nu}$ decays, which originate from $b \to u\ell\bar{\nu}$ quark-level processes, as can be seen in Fig. 8, and provide access to $|V_{ub}|$. If we complement this CKM matrix element with $|V_{cb}|$, we may determine the side R_b of the UT with the help of (36). The determination of $|V_{ub}|$ is hence a very important aspect of flavour physics. Since the π and ρ are 'light' mesons, the HQET symmetry relations cannot be applied to the $\bar{B} \to \pi\ell\bar{\nu}, \rho\ell\bar{\nu}$ modes. Consequently, in order to determine $|V_{ub}|$ from these exclusive channels, the corresponding heavy-to-light form factors have to be described by models. An important alternative is provided by inclusive decays. The corresponding decay rate takes the following form:

$$\Gamma(\bar{B} \to X_u\ell\bar{\nu}) = \frac{G_{\rm F}^2|V_{ub}|^2}{192\pi^3} m_b^5 \left[1 - 2.41\frac{\alpha_s}{\pi} + \frac{\lambda_1 - 9\lambda_2}{2m_b^2} + \ldots\right], \tag{73}$$

where λ_1 and λ_2 are non-perturbative parameters, which describe the hadronic matrix elements of certain 'kinetic' and 'chromomagnetic' operators appearing within the framework of the HQET. Using the heavy-quark expansions

$$M_B = m_b + \bar{\Lambda} - \frac{\lambda_1 + 3\lambda_2}{2m_b} + \ldots, \quad M_{B^*} = m_b + \bar{\Lambda} - \frac{\lambda_1 - \lambda_2}{2m_b} + \ldots \tag{74}$$

for the $B^{(*)}$-meson masses, where $\bar{\Lambda} \sim \Lambda_{\rm QCD}$ is another non-perturbative parameter that is related to the light degrees of freedom, the parameter λ_2 can be determined from the measured values of the $M_{B^{(*)}}$. The strong dependence of (73) on m_b is a significant source of uncertainty. On the other hand, the $1/m_b^2$ corrections can be better controlled than in the exclusive case (71), where we have, moreover, to deal with $1/m_c^2$ corrections. From an experimental point of view, we have to struggle with large backgrounds, which originate from $b \to c\ell\bar{\nu}$ processes and require also a model-dependent treatment. The determination of $|V_{ub}|$ from exclusive and inclusive B-meson decays caused by $b \to u\ell\bar{\nu}$ quark-level processes is therefore a very challenging issue; a summary of the current status is given by

$$|V_{ub}| = 0.0037 \times [1 \pm 0.15]. \tag{75}$$

If we now insert (75) and (72) into (36) and use $\lambda = 0.22$, we obtain

$$R_b = 0.41 \pm 0.07. \tag{76}$$

For a much more detailed discussion of the determinations of $|V_{cb}|$ and $|V_{ub}|$, addressing also the various interesting recent developments and future prospects, we refer the reader to Ref. [41], where the references to the vast original literature can also be found. Another excellent presentation is given in Ref. [17].

4.3 Non-leptonic decays

4.3.1 Classification

The most complicated B decays are the non-leptonic transitions, which are mediated by $b \to q_1\,\bar{q}_2\,d\,(s)$ quark-level processes, with $q_1, q_2 \in \{u, d, c, s\}$. There are two kinds of topologies contributing to such

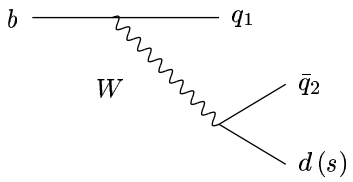


Fig. 9: Tree diagrams ($q_1, q_2 \in \{u, c\}$).

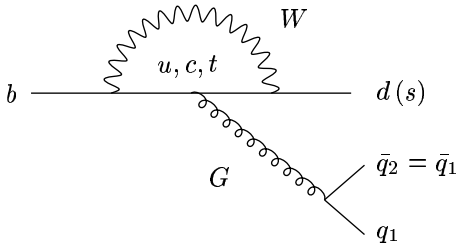
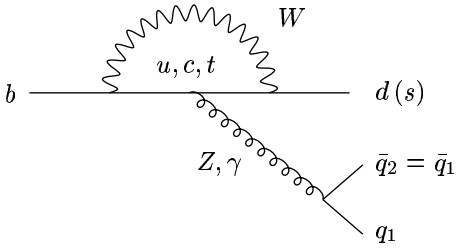


Fig. 10: QCD penguin diagrams ($q_1 = q_2 \in \{u, d, c, s\}$).



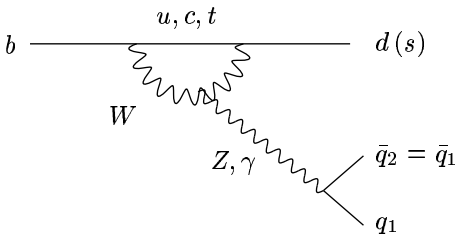


Fig. 11: Electroweak penguin diagrams ($q_1 = q_2 \in \{u, d, c, s\}$).

decays: tree-diagram-like and 'penguin' topologies. The latter consist of gluonic (QCD) and electroweak (EW) penguins. In Figs. 9–11, the corresponding leading-order Feynman diagrams are shown. Depending on the flavour content of their final states, we may classify $b \to q_1\,\bar{q}_2\,d\,(s)$ decays as follows:

- $q_1 \neq q_2 \in \{u, c\}$: *only* tree diagrams contribute.
- $q_1 = q_2 \in \{u, c\}$: tree *and* penguin diagrams contribute.
- $q_1 = q_2 \in \{d, s\}$: *only* penguin diagrams contribute.

4.3.2 Low-energy effective Hamiltonians

In order to analyse non-leptonic B decays theoretically, one uses low-energy effective Hamiltonians, which are calculated by making use of the 'operator product expansion', yielding transition matrix elements of the following structure:

$$\langle f|\mathcal{H}_{\rm eff}|i\rangle = \frac{G_{\rm F}}{\sqrt{2}}\lambda_{\rm CKM}\sum_k C_k(\mu)\langle f|Q_k(\mu)|i\rangle\,. \tag{77}$$

The technique of the operator product expansion allows us to separate the short-distance contributions to this transition amplitude from the long-distance ones, which are described by perturbative quantities $C_k(\mu)$ ('Wilson coefficient functions') and non-perturbative quantities $\langle f|Q_k(\mu)|i\rangle$ ('hadronic matrix elements'), respectively. As before, $G_{\rm F}$ is the Fermi constant, whereas $\lambda_{\rm CKM}$ is a CKM factor and μ denotes an appropriate renormalization scale. The Q_k are local operators, which are generated by electroweak interactions and QCD, and govern 'effectively' the decay in question. The Wilson coefficients $C_k(\mu)$ can be considered as scale-dependent couplings related to the vertices described by the Q_k.

In order to illustrate this rather abstract formalism, let us consider the decay $\bar{B}^0_d \to D^+K^-$, which allows a transparent discussion of the evaluation of the corresponding low-energy effective Hamiltonian. Since this transition originates from a $b \to c\bar{u}s$ quark-level process, it is—as we have seen in our classification in Subsection 4.3.1—a pure 'tree' decay, i.e. we do not have to deal with penguin topologies, which simplifies the analysis considerably. The leading-order Feynman diagram contributing to $\bar{B}^0_d \to D^+K^-$ can straightforwardly be obtained from Fig. 8 by substituting ℓ and ν by s and u, respectively. Consequently, the lepton current is simply replaced by a quark current, which will have important

implications shown below. Evaluating the corresponding Feynman diagram yields

$$-\frac{g_2^2}{8}V_{us}^*V_{cb}\left[\bar{s}\gamma^\nu(1-\gamma_5)u\right]\left[\frac{g_{\nu\mu}}{k^2-M_W^2}\right]\left[\bar{c}\gamma^\mu(1-\gamma_5)b\right]. \tag{78}$$

Because of $k^2 \sim m_b^2 \ll M_W^2$, we may—as in (58)—'integrate out' the W boson with the help of (53), and arrive at

$$\begin{aligned}\mathcal{H}_{\rm eff} &= \frac{G_{\rm F}}{\sqrt{2}}V_{us}^*V_{cb}\left[\bar{s}_\alpha\gamma_\mu(1-\gamma_5)u_\alpha\right]\left[\bar{c}_\beta\gamma^\mu(1-\gamma_5)b_\beta\right]\\ &= \frac{G_{\rm F}}{\sqrt{2}}V_{us}^*V_{cb}(\bar{s}_\alpha u_\alpha)_{\rm V-A}(\bar{c}_\beta b_\beta)_{\rm V-A} \equiv \frac{G_{\rm F}}{\sqrt{2}}V_{us}^*V_{cb}O_2\,,\end{aligned} \tag{79}$$

where α and β denote the colour indices of the $SU(3)_{\rm C}$ gauge group of QCD. Effectively, our $b \to c\bar{u}s$ decay process is now described by the 'current–current' operator O_2.

If we take QCD corrections into account, operator mixing induces a second 'current–current' operator, which is given by

$$O_1 \equiv \left[\bar{s}_\alpha\gamma_\mu(1-\gamma_5)u_\beta\right]\left[\bar{c}_\beta\gamma^\mu(1-\gamma_5)b_\alpha\right]. \tag{80}$$

Consequently, we obtain a low-energy effective Hamiltonian of the following structure:

$$\mathcal{H}_{\rm eff} = \frac{G_{\rm F}}{\sqrt{2}}V_{us}^*V_{cb}\left[C_1(\mu)O_1 + C_2(\mu)O_2\right], \tag{81}$$

where $C_1(\mu) \neq 0$ and $C_2(\mu) \neq 1$ are due to QCD renormalization effects [64]. In order to evaluate these coefficients, we must first calculate the QCD corrections to the decay processes both in the full theory, i.e. with W exchange, and in the effective theory, where the W is integrated out, and then express the QCD-corrected transition amplitude in terms of QCD-corrected matrix elements and Wilson coefficients as in (77). This procedure is called 'matching' between the full and the effective theory. The results for the $C_k(\mu)$ thus obtained contain terms of $\log(\mu/M_W)$, which become large for $\mu = \mathcal{O}(m_b)$, the scale governing the hadronic matrix elements of the O_k. Making use of the renormalization group, which exploits the fact that the transition amplitude (77) cannot depend on the chosen renormalization scale μ, we may sum up the following terms of the Wilson coefficients:

$$\alpha_s^n\left[\log\left(\frac{\mu}{M_W}\right)\right]^n \text{ (LO)}, \quad \alpha_s^n\left[\log\left(\frac{\mu}{M_W}\right)\right]^{n-1} \text{ (NLO)}, \quad \ldots \quad ; \tag{82}$$

detailed discussions of these rather technical aspects can be found in Refs. [22], [65].

For the exploration of CP violation, the class of non-leptonic B decays that receives contributions both from tree and from penguin topologies plays a key role. In this important case, the operator basis is much larger than in our example (81), where we considered a pure 'tree' decay. If we apply the relation

$$V_{ur}^*V_{ub} + V_{cr}^*V_{cb} + V_{tr}^*V_{tb} = 0 \quad (r \in \{d,s\}), \tag{83}$$

which follows from the unitarity of the CKM matrix, and 'integrate out' the top quark (which enters through the penguin loop processes) and the W boson, we may write

$$\mathcal{H}_{\rm eff} = \frac{G_{\rm F}}{\sqrt{2}}\left[\sum_{j=u,c}V_{jr}^*V_{jb}\left\{\sum_{k=1}^{2}C_k(\mu)\,Q_k^{jr} + \sum_{k=3}^{10}C_k(\mu)\,Q_k^{r}\right\}\right]. \tag{84}$$

Here we have introduced another quark-flavour label $j \in \{u,c\}$, and the Q_k^{jr} can be divided as follows:

– Current–current operators:

$$\begin{aligned} Q_1^{jr} &= (\bar{r}_\alpha j_\beta)_{\text{V–A}}(\bar{j}_\beta b_\alpha)_{\text{V–A}} \\ Q_2^{jr} &= (\bar{r}_\alpha j_\alpha)_{\text{V–A}}(\bar{j}_\beta b_\beta)_{\text{V–A}}. \end{aligned} \tag{85}$$

– QCD penguin operators:

$$\begin{aligned} Q_3^r &= (\bar{r}_\alpha b_\alpha)_{\text{V–A}} \textstyle\sum_{q'}(\bar{q}'_\beta q'_\beta)_{\text{V–A}} \\ Q_4^r &= (\bar{r}_\alpha b_\beta)_{\text{V–A}} \textstyle\sum_{q'}(\bar{q}'_\beta q'_\alpha)_{\text{V–A}} \\ Q_5^r &= (\bar{r}_\alpha b_\alpha)_{\text{V–A}} \textstyle\sum_{q'}(\bar{q}'_\beta q'_\beta)_{\text{V+A}} \\ Q_6^r &= (\bar{r}_\alpha b_\beta)_{\text{V–A}} \textstyle\sum_{q'}(\bar{q}'_\beta q'_\alpha)_{\text{V+A}}. \end{aligned} \tag{86}$$

– EW penguin operators (the $e_{q'}$ denote the electrical quark charges):

$$\begin{aligned} Q_7^r &= \tfrac{3}{2}(\bar{r}_\alpha b_\alpha)_{\text{V–A}} \textstyle\sum_{q'} e_{q'}(\bar{q}'_\beta q'_\beta)_{\text{V+A}} \\ Q_8^r &= \tfrac{3}{2}(\bar{r}_\alpha b_\beta)_{\text{V–A}} \textstyle\sum_{q'} e_{q'}(\bar{q}'_\beta q'_\alpha)_{\text{V+A}} \\ Q_9^r &= \tfrac{3}{2}(\bar{r}_\alpha b_\alpha)_{\text{V–A}} \textstyle\sum_{q'} e_{q'}(\bar{q}'_\beta q'_\beta)_{\text{V–A}} \\ Q_{10}^r &= \tfrac{3}{2}(\bar{r}_\alpha b_\beta)_{\text{V–A}} \textstyle\sum_{q'} e_{q'}(\bar{q}'_\beta q'_\alpha)_{\text{V–A}}. \end{aligned} \tag{87}$$

The current–current, QCD and EW penguin operators are related to the tree, QCD and EW penguin processes shown in Figs. 9–11. At a renormalization scale $\mu = \mathcal{O}(m_b)$, the Wilson coefficients of the current–current operators are $C_1(\mu) = \mathcal{O}(10^{-1})$ and $C_2(\mu) = \mathcal{O}(1)$, whereas those of the penguin operators are $\mathcal{O}(10^{-2})$ [22], [65]. Note that penguin topologies with internal charm- and up-quark exchanges [66] are described in this framework by penguin-like matrix elements of the corresponding current–current operators [67], and may also have important phenomenological consequences [68], [69].

Since the ratio $\alpha/\alpha_s = \mathcal{O}(10^{-2})$ of the QED and QCD couplings is very small, we would expect naïvely that EW penguins should play a minor role in comparison with QCD penguins. This would actually be the case if the top quark was not 'heavy'. However, since the Wilson coefficient C_9 increases strongly with m_t, we obtain interesting EW penguin effects in several B decays: $B \to K\phi$ modes are affected significantly by EW penguins, whereas $B \to \pi\phi$ and $B_s \to \pi^0\phi$ transitions are even *dominated* by such topologies [70], [71]. EW penguins also have an important impact on the $B \to \pi K$ system [72], as we shall see in Subsection 7.2.

The low-energy effective Hamiltonians discussed above apply to all B decays that are caused by the same quark-level transition, i.e. they are 'universal'. Consequently, the differences between the various exclusive modes of a given decay class arise within this formalism only through the hadronic matrix elements of the relevant four-quark operators. Unfortunately, the evaluation of such matrix elements is associated with large uncertainties and is a very challenging task. In this context, 'factorization' is a widely used concept, which is our next topic.

4.3.3 Factorization of hadronic matrix elements

In order to discuss 'factorization', let us consider once more the decay $\bar{B}_d^0 \to D^+K^-$. Evaluating the corresponding transition amplitude, we encounter the hadronic matrix elements of the $O_{1,2}$ operators between the $\langle K^-D^+|$ final and the $|\bar{B}_d^0\rangle$ initial states. If we use the well-known $SU(N_{\text{C}})$ colour-algebra relation

$$T^a_{\alpha\beta}T^a_{\gamma\delta} = \frac{1}{2}\left(\delta_{\alpha\delta}\delta_{\beta\gamma} - \frac{1}{N_{\text{C}}}\delta_{\alpha\beta}\delta_{\gamma\delta}\right) \tag{88}$$

to rewrite the operator O_1, we obtain

$$\begin{aligned} \langle K^-D^+|\mathcal{H}_{\text{eff}}|\bar{B}_d^0\rangle = \frac{G_{\text{F}}}{\sqrt{2}}V_{us}^*V_{cb}\Big[a_1\langle K^-D^+|(\bar{s}_\alpha u_\alpha)_{\text{V–A}}(\bar{c}_\beta b_\beta)_{\text{V–A}}|\bar{B}_d^0\rangle \\ +2\,C_1\langle K^-D^+|(\bar{s}_\alpha\, T^a_{\alpha\beta}\, u_\beta)_{\text{V–A}}(\bar{c}_\gamma\, T^a_{\gamma\delta}\, b_\delta)_{\text{V–A}}|\bar{B}_d^0\rangle\Big], \end{aligned}$$

with

$$a_1 = C_1/N_{\text{C}} + C_2 \sim 1. \tag{89}$$

It is now straightforward to 'factorize' the hadronic matrix elements in (89):

$$\begin{aligned}
&\langle K^- D^+ | (\bar{s}_\alpha u_\alpha)_{\text{V–A}} (\bar{c}_\beta b_\beta)_{\text{V–A}} | \bar{B}_d^0 \rangle \big|_{\text{fact}} \\
&\quad = \langle K^- | \left[\bar{s}_\alpha \gamma_\mu (1-\gamma_5) u_\alpha \right] | 0 \rangle \langle D^+ | \left[\bar{c}_\beta \gamma^\mu (1-\gamma_5) b_\beta \right] | \bar{B}_d^0 \rangle \\
&\quad = \underbrace{i f_K}_{\text{decay constant}} \times \underbrace{F_0^{(BD)}(M_K^2)}_{B \to D \text{ form factor}} \times \underbrace{(M_B^2 - M_D^2),}_{\text{kinematical factor}}
\end{aligned} \tag{90}$$

$$\langle K^- D^+ | (\bar{s}_\alpha \, T^a_{\alpha\beta} \, u_\beta)_{\text{V–A}} (\bar{c}_\gamma \, T^a_{\gamma\delta} \, b_\delta)_{\text{V–A}} | \bar{B}_d^0 \rangle \big|_{\text{fact}} = 0. \tag{91}$$

The quantity a_1 is a phenomenological 'colour factor', which governs 'colour-allowed' decays; the decay $\bar{B}_d^0 \to D^+ K^-$ belongs to this category, since the colour indices of the K^- meson and the $\bar{B}_d^0$–D^+ system run independently from each other in the corresponding leading-order diagram. On the other hand, in the case of 'colour-suppressed' modes, for instance $\bar{B}_d^0 \to \pi^0 D^0$, where only one colour index runs through the whole diagram, we have to deal with the combination

$$a_2 = C_1 + C_2/N_{\text{C}} \sim 0.25. \tag{92}$$

The concept of factorizing the hadronic matrix elements of four-quark operators into the product of hadronic matrix elements of quark currents has a long history [73], and can be justified, for example, in the large-N_{C} limit [74]. Interesting recent developments are the following:

– 'QCD factorization' [75], which is in accordance with the old picture that factorization should hold for certain decays in the limit of $m_b \gg \Lambda_{\text{QCD}}$ [76], provides a formalism to calculate the relevant amplitudes at the leading order of a Λ_{QCD}/m_b expansion. The resulting expression for the transition amplitudes incorporates elements both of the naïve factorization approach sketched above and of the hard-scattering picture. Let us consider a decay $\bar{B} \to M_1 M_2$, where M_1 picks up the spectator quark. If M_1 is either a heavy (D) or a light (π, K) meson, and M_2 a light (π, K) meson, QCD factorization gives a transition amplitude of the following structure:

$$A(\bar{B} \to M_1 M_2) = [\text{'naïve factorization'}] \times [1 + \mathcal{O}(\alpha_s) + \mathcal{O}(\Lambda_{\text{QCD}}/m_b)] \,. \tag{93}$$

  While the $\mathcal{O}(\alpha_s)$ terms, i.e. the radiative non-factorizable corrections, can be calculated systematically, the main limitation of the theoretical accuracy originates from the $\mathcal{O}(\Lambda_{\text{QCD}}/\, m_b)$ terms.

– Another QCD approach to deal with non-leptonic B-meson decays — PQCD, the 'perturbative hard-scattering approach' — was developed independently in Ref. [77], and differs from the QCD factorization formalism in some technical aspects.

– A very useful technique for 'factorization proofs' is provided by the framework of the 'soft collinear effective theory' (SCET) [78].

– Non-leptonic B decays can also be studied within QCD light-cone sum-rule approaches [79].

A detailed presentation of these topics would be very technical and is beyond the scope of these lectures. However, for the discussion of the CP-violating effects in the B-meson system, we must only be familiar with the general structure of the non-leptonic B decay amplitudes and not enter the details of the techniques to deal with the corresponding hadronic matrix elements. Let us finally note that the B-factory data will eventually decide how well factorization and the new concepts sketched above are actually working. For example, recent data on the $B \to \pi\pi$ system point towards large non-factorizable corrections [48], [54], to which we shall return in Subsection 6.2.2.

4.4 Towards studies of CP violation

As we have seen above, leptonic and semileptonic B-meson decays involve only a single weak (CKM) amplitude. On the other hand, the structure of non-leptonic transitions is considerably more complicated.

However, because of the unitarity of the CKM matrix, which implies the relation in (83), we may write the amplitude of *any* non-leptonic B-meson decay within the SM in such a manner that we encounter at most two contributions with different CKM factors (we shall encounter explicit examples below):

$$A(\bar{B} \to \bar{f}) = e^{+i\varphi_1}|A_1|e^{i\delta_1} + e^{+i\varphi_2}|A_2|e^{i\delta_2} \tag{94}$$

$$A(B \to f) = e^{-i\varphi_1}|A_1|e^{i\delta_1} + e^{-i\varphi_2}|A_2|e^{i\delta_2}. \tag{95}$$

Here the $\varphi_{1,2}$ denote CP-violating weak phases, which are introduced by the elements of the CKM matrix, whereas the $|A_{1,2}|e^{i\delta_{1,2}}$ are CP-conserving 'strong' amplitudes, which contain the whole hadron dynamics of the decay at hand:

$$|A|e^{i\delta} \sim \sum_k \underbrace{C_k(\mu)}_{\text{pert. QCD}} \times \underbrace{\langle \bar{f}|Q_k(\mu)|\bar{B}\rangle}_{\text{non-pert. QCD}}. \tag{96}$$

If we use (94) and (95), it is an easy exercise to calculate the following CP-violating rate asymmetry:

$$\begin{aligned}\mathcal{A}_{\rm CP} &\equiv \frac{\Gamma(B \to f) - \Gamma(\bar{B} \to \bar{f})}{\Gamma(B \to f) + \Gamma(\bar{B} \to \bar{f})} = \frac{|A(B \to f)|^2 - |A(\bar{B} \to \bar{f})|^2}{|A(B \to f)|^2 + |A(\bar{B} \to \bar{f})|^2} \\ &= \frac{2|A_1||A_2|\sin(\delta_1 - \delta_2)\sin(\varphi_1 - \varphi_2)}{|A_1|^2 + 2|A_1||A_2|\cos(\delta_1 - \delta_2)\cos(\varphi_1 - \varphi_2) + |A_2|^2}.\end{aligned} \tag{97}$$

Consequently, a non-vanishing CP asymmetry $\mathcal{A}_{\rm CP}$ arises from the interference effects between the two weak amplitudes, and requires both a non-trivial weak phase difference $\varphi_1 - \varphi_2$ and a non-trivial strong phase difference $\delta_1 - \delta_2$. This kind of CP violation is referred to as 'direct' CP violation, as it originates directly at the amplitude level of the considered decay. It is the B-meson counterpart of the effects that are probed through $\mathrm{Re}(\varepsilon'/\varepsilon)$ in the neutral kaon system.[2] Since $\varphi_1 - \varphi_2$ is in general given by one of the angles of the UT—usually γ—the goal is to determine this quantity from the measured value of $\mathcal{A}_{\rm CP}$. Unfortunately, the extraction of $\varphi_1 - \varphi_2$ from $\mathcal{A}_{\rm CP}$ is affected by hadronic uncertainties, which are related to the poorly known hadronic matrix elements entering the expression (96) for the strong amplitudes $|A_{1,2}|e^{i\delta_{1,2}}$. In order to deal with this problem, we may, in principle, proceed along one of the following three main avenues:

i) The most obvious one—but also the most challenging—is to try to *calculate* the relevant hadronic matrix elements $\langle \bar{f}|Q_k(\mu)|\bar{B}\rangle$. As we have noted above, interesting progress has recently been made in this direction through the development of the QCD factorization, PQCD, SCET and QCD light-cone sum-rule approaches.

ii) We may search for fortunate cases, where relations between various decay amplitudes allow us to *eliminate* the poorly known hadronic matrix elements. As we shall see, this avenue offers in particular determinations of the UT angle γ: we distinguish between exact relations, which are provided by pure 'tree' decays of the kind $B \to KD$ or $B_c \to D_s D$, and relations, which follow from the flavour symmetries of strong interactions, involving $B_{(s)} \to \pi\pi, \pi K, KK$ transitions.

iii) Finally, we may exploit the fact that in decays of neutral B_q mesons ($q \in \{d, s\}$) interference effects between B_q^0–$\bar{B}_q^0$ mixing and decay processes may yield another kind of CP violation, 'mixing-induced CP violation'. In certain cases, the hadronic matrix elements *cancel* in such CP asymmeties.

In the remainder of these lectures, we shall not consider (i) further. For the exploration of CP violation and the testing of the KM mechanism, the theoretical input related to strong-interaction physics should obviously be reduced as much as possible. In contrast to (i), this feature is present in (ii) and (iii),

[2] In order to calculate this quantity, an appropriate low-energy effective Hamiltonian having the same structure as (84) is used. The large theoretical uncertainties mentioned in Subsection 3.1 originate from a strong cancellation between the contributions of the QCD and EW penguins (caused by the large top-quark mass) and the associated hadronic matrix elements.

which provide—as a by-product—also important insights into hadron dynamics. In particular, we may extract various hadronic parameters from the data that can be calculated with the help of the theoretical frameworks listed in (i), thereby allowing us to test them through a confrontation with nature. Since neutral B_q mesons are a key element in this programme, offering also attractive connections between (ii) and (iii), let us next have a closer look at their most important features.

5 FEATURES OF NEUTRAL $B_{d,s}$ MESONS

5.1 $B^0_{d,s}$–$\bar{B}^0_{d,s}$ mixing

Within the SM, B^0_q–$\bar{B}^0_q$ mixing ($q \in \{d, s\}$) arises from the box diagrams shown in Fig. 12. Because of this phenomenon, an initially, i.e. at time $t = 0$, present B^0_q-meson state evolves into a time-dependent linear combination of B^0_q and $\bar{B}^0_q$ states:

$$|B_q(t)\rangle = a(t)|B^0_q\rangle + b(t)|\bar{B}^0_q\rangle, \tag{98}$$

where $a(t)$ and $b(t)$ are governed by a Schrödinger equation of the following form:

$$i\,\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = H \cdot \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} \equiv \left[\underbrace{\begin{pmatrix} M_0^{(q)} & M_{12}^{(q)} \\ M_{12}^{(q)*} & M_0^{(q)} \end{pmatrix}}_{\text{mass matrix}} - \frac{i}{2}\underbrace{\begin{pmatrix} \Gamma_0^{(q)} & \Gamma_{12}^{(q)} \\ \Gamma_{12}^{(q)*} & \Gamma_0^{(q)} \end{pmatrix}}_{\text{decay matrix}}\right] \cdot \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}.$$

The special form $H_{11} = H_{22}$ of the Hamiltonian H is an implication of the CPT theorem, i.e. of the invariance under combined CP and time-reversal (T) transformations.

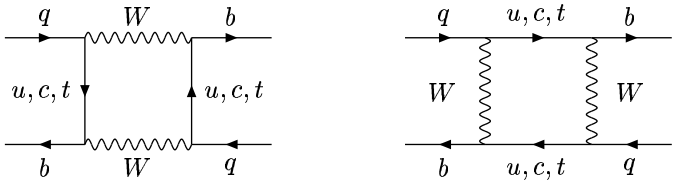


Fig. 12: Box diagrams contributing to B^0_q–$\bar{B}^0_q$ mixing in the SM ($q \in \{d, s\}$).

5.1.1 Solution of the Schrödinger equation

It is straightforward to calculate the eigenstates $|B^{(q)}_\pm\rangle$ and eigenvalues $\lambda^{(q)}_\pm$ of (99):

$$|B^{(q)}_\pm\rangle = \frac{1}{\sqrt{1 + |\alpha_q|^2}}\left(|B^0_q\rangle \pm \alpha_q|\bar{B}^0_q\rangle\right) \tag{99}$$

$$\lambda^{(q)}_\pm = \left(M_0^{(q)} - \frac{i}{2}\Gamma_0^{(q)}\right) \pm \left(M_{12}^{(q)} - \frac{i}{2}\Gamma_{12}^{(q)}\right)\alpha_q, \tag{100}$$

where

$$\alpha_q e^{+i\left(\Theta^{(q)}_{\Gamma_{12}} + n'\pi\right)} = \sqrt{\frac{4|M_{12}^{(q)}|^2 e^{-i2\delta\Theta^{(q)}_{M/\Gamma}} + |\Gamma_{12}^{(q)}|^2}{4|M_{12}^{(q)}|^2 + |\Gamma_{12}^{(q)}|^2 - 4|M_{12}^{(q)}||\Gamma_{12}^{(q)}|\sin\delta\Theta^{(q)}_{M/\Gamma}}}. \tag{101}$$

Here we have written

$$M_{12}^{(q)} \equiv e^{i\Theta^{(q)}_{M_{12}}}|M_{12}^{(q)}|, \quad \Gamma_{12}^{(q)} \equiv e^{i\Theta^{(q)}_{\Gamma_{12}}}|\Gamma_{12}^{(q)}|, \quad \delta\Theta^{(q)}_{M/\Gamma} \equiv \Theta^{(q)}_{M_{12}} - \Theta^{(q)}_{\Gamma_{12}}, \tag{102}$$

and have introduced the quantity $n' \in \{0, 1\}$ to parametrize the sign of the square root in (101).

Evaluating the dispersive parts of the box diagrams shown in Fig 12, which are dominated by internal top-quark exchanges, yields (for a more detailed discussion, see [21]):

$$M_{12}^{(q)} = \frac{G_{\rm F}^2 M_W^2}{12\pi^2}\eta_B M_{B_q}\hat{B}_{B_q} f_{B_q}^2 \left(V_{tq}^* V_{tb}\right)^2 S_0(x_t) e^{i(\pi-\phi_{\rm CP}(B_q))}, \tag{103}$$

where $\eta_B = 0.55 \pm 0.01$ is a perturbative QCD correction [80],[3] the non-perturbative 'bag' parameter $\hat{B}_{B_q}$ is related to the hadronic matrix element $\langle \bar{B}_q^0|(\bar{b}q)_{\rm V-A}(\bar{b}q)_{\rm V-A}|B_q^0\rangle$, and $S_0(x_t \equiv m_t^2/M_W^2)$ is one of the 'Inami–Lim' functions [81], describing the dependence on the top-quark mass m_t. In the SM, we may write—to a good approximation—the following expression [82]:

$$S_0(x_t) = 2.40 \times \left[\frac{m_t}{167\,\text{GeV}}\right]^{1.52}. \tag{104}$$

Finally, $\phi_{\rm CP}(B_q)$ is a convention-dependent phase, which is introduced through the CP transformation

$$(\mathcal{CP})|B_q^0\rangle = e^{i\phi_{\rm CP}(B_q)}|\bar{B}_q^0\rangle. \tag{105}$$

If we calculate also the absorptive parts of the box diagrams in Fig 12, we obtain

$$\frac{\Gamma_{12}^{(q)}}{M_{12}^{(q)}} \approx -\frac{3\pi}{2S_0(x_t)}\left(\frac{m_b^2}{M_W^2}\right) = \mathcal{O}(m_b^2/m_t^2) \ll 1. \tag{106}$$

Consequently, we may expand (101) in $\Gamma_{12}^{(q)}/M_{12}^{(q)}$. Neglecting second-order terms, we arrive at

$$\alpha_q = \left[1 + \frac{1}{2}\left|\frac{\Gamma_{12}^{(q)}}{M_{12}^{(q)}}\right| \sin\delta\Theta_{M/\Gamma}^{(q)}\right] e^{-i\left(\Theta_{M_{12}}^{(q)} + n'\pi\right)}. \tag{107}$$

The deviation of $|\alpha_q|$ from 1 measures CP violation in B_q^0–$\bar{B}_q^0$ oscillations, and can be probed through the following 'wrong-charge' lepton asymmetries:

$$\mathcal{A}_{\rm SL}^{(q)} \equiv \frac{\Gamma(B_q^0(t) \to \ell^-\bar{\nu}X) - \Gamma(\bar{B}_q^0(t) \to \ell^+\nu X)}{\Gamma(B_q^0(t) \to \ell^-\bar{\nu}X) + \Gamma(\bar{B}_q^0(t) \to \ell^+\nu X)} = \frac{|\alpha_q|^4 - 1}{|\alpha_q|^4 + 1} \approx \left|\frac{\Gamma_{12}^{(q)}}{M_{12}^{(q)}}\right| \sin\delta\Theta_{M/\Gamma}^{(q)}. \tag{108}$$

Because of $|\Gamma_{12}^{(q)}|/|M_{12}^{(q)}| \propto m_b^2/m_t^2$ and $\sin\delta\Theta_{M/\Gamma}^{(q)} \propto m_c^2/m_b^2$, the asymmetry $\mathcal{A}_{\rm SL}^{(q)}$ is suppressed by a factor of $m_c^2/m_t^2 = \mathcal{O}(10^{-4})$ and is hence tiny in the SM. However, this observable may be enhanced through NP effects, thereby representing an interesing probe for physics beyond the SM [83], [84]. The current experimental constraints for $\mathcal{A}_{\rm SL}^{(q)}$ are at the 10^{-2} level.

5.1.2 Mixing parameters

Let us denote the masses of the eigenstates of (99) by $M_{\rm H}^{(q)}$ ('heavy') and $M_{\rm L}^{(q)}$ ('light'). It is then useful to introduce

$$M_q \equiv \frac{M_{\rm H}^{(q)} + M_{\rm L}^{(q)}}{2} = M_0^{(q)}, \tag{109}$$

as well as the mass difference

$$\Delta M_q \equiv M_{\rm H}^{(q)} - M_{\rm L}^{(q)} = 2|M_{12}^{(q)}| > 0, \tag{110}$$

[3] Note that the short-distance parameter η_B does *not* depend on $q \in \{d, s\}$, i.e. is the same for B_d and B_s mesons.

which is by definition *positive*. Using (37) and (103), we find that we may convert the mass difference ΔM_d of the B_d-meson system into the side R_t of the UT with the help of the following expression:

$$R_t = \frac{1.10}{A\sqrt{|S_0(x_t)|}}\sqrt{\frac{\Delta M_d}{0.50\,\mathrm{ps}^{-1}}}\left[\frac{230\,\mathrm{MeV}}{\sqrt{\hat{B}_{B_d}}f_{B_d}}\right]\sqrt{\frac{0.55}{\eta_B}}, \tag{111}$$

where A is the usual Wolfenstein parameter. We shall return to this important issue in Subsection 8.1.2.

On the other hand, the decay widths $\Gamma_{\rm H}^{(q)}$ and $\Gamma_{\rm L}^{(q)}$ of the mass eigenstates, which correspond to $M_{\rm H}^{(q)}$ and $M_{\rm L}^{(q)}$, respectively, satisfy

$$\Delta\Gamma_q \equiv \Gamma_{\rm H}^{(q)} - \Gamma_{\rm L}^{(q)} = \frac{4\,\mathrm{Re}\left[M_{12}^{(q)}\Gamma_{12}^{(q)*}\right]}{\Delta M_q}, \tag{112}$$

whereas

$$\Gamma_q \equiv \frac{\Gamma_{\rm H}^{(q)} + \Gamma_{\rm L}^{(q)}}{2} = \Gamma_0^{(q)}. \tag{113}$$

There is the following interesting relation:

$$\frac{\Delta\Gamma_q}{\Gamma_q} \approx -\frac{3\pi}{2S_0(x_t)}\left(\frac{m_b^2}{M_W^2}\right)x_q = -\mathcal{O}(10^{-2})\times x_q, \tag{114}$$

where

$$x_q \equiv \frac{\Delta M_q}{\Gamma_q} = \begin{cases} 0.771 \pm 0.012 & (q=d) \\ \mathcal{O}(20) & (q=s) \end{cases} \tag{115}$$

denotes the B_q^0–$\bar{B}_q^0$ 'mixing parameter'.[4] Consequently, we observe that $\Delta\Gamma_d/\Gamma_d \sim 10^{-2}$ is negligibly small, while $\Delta\Gamma_s/\Gamma_s \sim 10^{-1}$ may be sizeable. For a discussion of the experimental status of the B_q mixing parameters, the reader is referred to Refs. [85], [86].

5.1.3 Time-dependent decay rates

The time evolution of initially, i.e. at $t=0$, pure B_q^0- and $\bar{B}_q^0$-meson states is given by

$$|B_q^0(t)\rangle = f_+^{(q)}(t)|B_q^0\rangle + \alpha_q f_-^{(q)}(t)|\bar{B}_q^0\rangle \tag{116}$$

and

$$|\bar{B}_q^0(t)\rangle = \frac{1}{\alpha_q}f_-^{(q)}(t)|B_q^0\rangle + f_+^{(q)}(t)|\bar{B}_q^0\rangle, \tag{117}$$

respectively, with

$$f_\pm^{(q)}(t) = \frac{1}{2}\left[e^{-i\lambda_+^{(q)}t} \pm e^{-i\lambda_-^{(q)}t}\right]. \tag{118}$$

These time-dependent state vectors allow the calculation of the corresponding transition rates. To this end, it is useful to introduce

$$|g_\pm^{(q)}(t)|^2 = \frac{1}{4}\left[e^{-\Gamma_{\rm L}^{(q)}t} + e^{-\Gamma_{\rm H}^{(q)}t} \pm 2\,e^{-\Gamma_q t}\cos(\Delta M_q t)\right] \tag{119}$$

$$g_-^{(q)}(t)\,g_+^{(q)}(t)^* = \frac{1}{4}\left[e^{-\Gamma_{\rm L}^{(q)}t} - e^{-\Gamma_{\rm H}^{(q)}t} + 2\,i\,e^{-\Gamma_q t}\sin(\Delta M_q t)\right], \tag{120}$$

as well as

$$\xi_f^{(q)} = e^{-i\Theta_{M_{12}}^{(q)}}\frac{A(\bar{B}_q^0\to f)}{A(B_q^0\to f)}, \quad \xi_{\bar{f}}^{(q)} = e^{-i\Theta_{M_{12}}^{(q)}}\frac{A(\bar{B}_q^0\to \bar{f})}{A(B_q^0\to \bar{f})}. \tag{121}$$

[4] Note that $\Delta\Gamma_q/\Gamma_q$ is negative in the SM because of the minus sign in (114).

Looking at (103), we find

$$\Theta_{M_{12}}^{(q)} = \pi + 2\text{arg}(V_{tq}^* V_{tb}) - \phi_{\text{CP}}(B_q), \tag{122}$$

and observe that this phase depends on the chosen CKM and CP phase conventions specified in (9) and (105), respectively. However, these dependences are cancelled through the amplitude ratios in (121), so that $\xi_f^{(q)}$ and $\xi_{\bar{f}}^{(q)}$ are *convention-independent* observables. Whereas n' enters the functions in (118) through (100), the dependence on this parameter is cancelled in (119) and (120) through the introduction of the *positive* mass difference ΔM_q [see (110)]. Combining the formulae listed above, we eventually arrive at the following transition rates for decays of initially, i.e. at $t = 0$, present B_q^0 or $\bar{B}_q^0$ mesons:

$$\Gamma(\overset{(-)}{B}{}_q^0\,(t) \to f) = \left[|g_{\mp}^{(q)}(t)|^2 + |\xi_f^{(q)}|^2 |g_{\pm}^{(q)}(t)|^2 - 2\,\text{Re}\left\{\xi_f^{(q)} g_{\pm}^{(q)}(t) g_{\mp}^{(q)}(t)^*\right\}\right] \tilde{\Gamma}_f, \tag{123}$$

where the time-independent rate $\tilde{\Gamma}_f$ corresponds to the ' unevolved' decay amplitude $A(B_q^0 \to f)$, and can be calculated by performing the usual phase-space integrations. The rates into the CP-conjugate final state $\bar{f}$ can straightforwardly be obtained from (123) by making the substitutions

$$\tilde{\Gamma}_f \;\to\; \tilde{\Gamma}_{\bar{f}}, \qquad \xi_f^{(q)} \;\to\; \xi_{\bar{f}}^{(q)}. \tag{124}$$

5.2 CP asymmetries

A particularly simple—but also very interesting—situation arises if we restrict ourselves to decays of neutral B_q mesons into final states f that are eigenstates of the CP operator, i.e. satisfy the relation

$$(\mathcal{CP})|f\rangle = \pm|f\rangle. \tag{125}$$

Consequently, we have $\xi_f^{(q)} = \xi_{\bar{f}}^{(q)}$ in this case, as can be seen in (121). Using the decay rates in (123), we find that the corresponding time-dependent CP asymmetry is given by

$$\begin{aligned} \mathcal{A}_{\text{CP}}(t) &\equiv \frac{\Gamma(B_q^0(t) \to f) - \Gamma(\bar{B}_q^0(t) \to f)}{\Gamma(B_q^0(t) \to f) + \Gamma(\bar{B}_q^0(t) \to f)} \\ &= \left[\frac{\mathcal{A}_{\text{CP}}^{\text{dir}}(B_q \to f)\,\cos(\Delta M_q t) + \mathcal{A}_{\text{CP}}^{\text{mix}}(B_q \to f)\,\sin(\Delta M_q t)}{\cosh(\Delta\Gamma_q t/2) - \mathcal{A}_{\Delta\Gamma}(B_q \to f)\,\sinh(\Delta\Gamma_q t/2)}\right], \end{aligned} \tag{126}$$

with

$$\mathcal{A}_{\text{CP}}^{\text{dir}}(B_q \to f) \equiv \frac{1 - \left|\xi_f^{(q)}\right|^2}{1 + \left|\xi_f^{(q)}\right|^2}, \qquad \mathcal{A}_{\text{CP}}^{\text{mix}}(B_q \to f) \equiv \frac{2\,\text{Im}\,\xi_f^{(q)}}{1 + \left|\xi_f^{(q)}\right|^2}. \tag{127}$$

Because of the relation

$$\mathcal{A}_{\text{CP}}^{\text{dir}}(B_q \to f) = \frac{|A(B_q^0 \to f)|^2 - |A(\bar{B}_q^0 \to \bar{f})|^2}{|A(B_q^0 \to f)|^2 + |A(\bar{B}_q^0 \to \bar{f})|^2}, \tag{128}$$

this observable measures the direct CP violation in the decay $B_q \to f$, which originates from the interference between different weak amplitudes, as we have seen in (97). On the other hand, the interesting *new* aspect of (126) is due to $\mathcal{A}_{\text{CP}}^{\text{mix}}(B_q \to f)$, which originates from interference effects between B_q^0–$\bar{B}_q^0$ mixing and decay processes, and describes 'mixing-induced' CP violation. Finally, the width difference $\Delta\Gamma_q$, which may be sizeable in the B_s-meson system, provides another observable,

$$\mathcal{A}_{\Delta\Gamma}(B_q \to f) \equiv \frac{2\,\text{Re}\,\xi_f^{(q)}}{1 + \left|\xi_f^{(q)}\right|^2}, \tag{129}$$

which is, however, not independent from $\mathcal{A}_{\rm CP}^{\rm dir}(B_q \to f)$ and $\mathcal{A}_{\rm CP}^{\rm mix}(B_q \to f)$, satisfying

$$\left[\mathcal{A}_{\rm CP}^{\rm dir}(B_q \to f)\right]^2 + \left[\mathcal{A}_{\rm CP}^{\rm mix}(B_q \to f)\right]^2 + \left[\mathcal{A}_{\Delta\Gamma}(B_q \to f)\right]^2 = 1. \tag{130}$$

In order to calculate the quantity $\xi_f^{(q)}$, which contains essentially all the information that is required for the evaluation of the observables provided by the time-dependent CP asymmetry introduced in (126), we employ the low-energy effective Hamiltonian (84):

$$\begin{aligned} A(\bar{B}_q^0 \to f) &= \langle f|\mathcal{H}_{\rm eff}|\bar{B}_q^0\rangle \\ &= \frac{G_{\rm F}}{\sqrt{2}}\left[\sum_{j=u,c} V_{jr}^* V_{jb}\left\{\sum_{k=1}^{2} C_k(\mu)\langle f|Q_k^{jr}(\mu)|\bar{B}_q^0\rangle + \sum_{k=3}^{10} C_k(\mu)\langle f|Q_k^{r}(\mu)|\bar{B}_q^0\rangle\right\}\right]. \end{aligned} \tag{131}$$

On the other hand, we also have

$$\begin{aligned} A(B_q^0 \to f) &= \langle f|\mathcal{H}_{\rm eff}^\dagger|B_q^0\rangle \\ &= \frac{G_{\rm F}}{\sqrt{2}}\left[\sum_{j=u,c} V_{jr} V_{jb}^*\left\{\sum_{k=1}^{2} C_k(\mu)\langle f|Q_k^{jr\dagger}(\mu)|B_q^0\rangle + \sum_{k=3}^{10} C_k(\mu)\langle f|Q_k^{r\dagger}(\mu)|B_q^0\rangle\right\}\right]. \end{aligned} \tag{132}$$

If we now insert the operator $(\mathcal{CP})^\dagger(\mathcal{CP}) = \hat{1}$ both after the $\langle f|$ and in front of the $|B_q^0\rangle$, we obtain

$$\begin{aligned} A(B_q^0 \to f) &= \pm e^{i\phi_{\rm CP}(B_q)} \\ &\times \frac{G_{\rm F}}{\sqrt{2}}\left[\sum_{j=u,c} V_{jr} V_{jb}^*\left\{\sum_{k=1}^{2} C_k(\mu)\langle f|Q_k^{jr}(\mu)|\bar{B}_q^0\rangle + \sum_{k=3}^{10} C_k(\mu)\langle f|Q_k^{r}(\mu)|\bar{B}_q^0\rangle\right\}\right], \end{aligned} \tag{133}$$

where we have also applied the relation $(\mathcal{CP})Q_k^{jr\dagger}(\mathcal{CP})^\dagger = Q_k^{jr}$, and have furthermore taken (105) into account. Using then (121) and (122), we observe that the phase-convention-dependent quantity $\phi_{\rm CP}(B_q)$ cancels, and finally arrive at

$$\xi_f^{(q)} = \mp\, e^{-i\phi_q}\left[\frac{\sum_{j=u,c} V_{jr}^* V_{jb}\langle f|\,\mathcal{Q}^{jr}\,|\bar{B}_q^0\rangle}{\sum_{j=u,c} V_{jr} V_{jb}^*\langle f|\,\mathcal{Q}^{jr}\,|\bar{B}_q^0\rangle}\right]. \tag{134}$$

Here we have introduced the abbreviation

$$\mathcal{Q}^{jr} \equiv \sum_{k=1}^{2} C_k(\mu)\, Q_k^{jr} + \sum_{k=3}^{10} C_k(\mu)\, Q_k^{r}, \tag{135}$$

and

$$\phi_q \equiv 2\arg(V_{tq}^* V_{tb}) = \begin{cases} +2\beta & (q = d) \\ -2\delta\gamma & (q = s) \end{cases} \tag{136}$$

(where β and $\delta\gamma$ are the angles in the unitarity triangles illustrated in Fig. 3) is the CP-violating weak phase introduced by B_q^0–$\bar{B}_q^0$ mixing within the SM.

Using the notation of (94) and (95), we may rewrite (134) as follows:

$$\xi_f^{(q)} = \mp\, e^{-i\phi_q}\left[\frac{e^{+i\varphi_1}|A_1|e^{i\delta_1} + e^{+i\varphi_2}|A_2|e^{i\delta_2}}{e^{-i\varphi_1}|A_1|e^{i\delta_1} + e^{-i\varphi_2}|A_2|e^{i\delta_2}}\right]. \tag{137}$$

In analogy to the discussion of direct CP violation in Subsection 4.4, the calculation of $\xi_f^{(q)}$ suffers—in general—from large hadronic uncertainties. However, if one CKM amplitude plays the *dominant* role in the transition $B_q \to f$, we obtain

$$\xi_f^{(q)} = \mp\, e^{-i\phi_q} \left[\frac{e^{+i\phi_f/2}|M_f|e^{i\delta_f}}{e^{-i\phi_f/2}|M_f|e^{i\delta_f}}\right] = \mp\, e^{-i(\phi_q-\phi_f)}, \tag{138}$$

and observe that the hadronic matrix element $|M_f|e^{i\delta_f}$ *cancels* in this expression. Since the requirements for direct CP violation discussed in the context of (97) are no longer satisfied, we have vanishing direct CP violation in this important special case, i.e. $\mathcal{A}_{\rm CP}^{\rm dir}(B_q \to f) = 0$, which is also obvious from (127) and (138). On the other hand, we still have mixing-induced CP violation. In particular,

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_q \to f) = \pm \sin\phi \tag{139}$$

is now governed by the CP-violating weak phase difference $\phi \equiv \phi_q - \phi_f$ and is *not* affected by hadronic uncertainties. The corresponding time-dependent CP asymmetry then takes the simple form

$$\left.\frac{\Gamma(B_q^0(t) \to f) - \Gamma(\bar{B}_q^0(t) \to \bar{f})}{\Gamma(B_q^0(t) \to f) + \Gamma(\bar{B}_q^0(t) \to \bar{f})}\right|_{\Delta\Gamma_q=0} = \pm \sin\phi\, \sin(\Delta M_q t), \tag{140}$$

and allows an elegant determination of $\sin\phi$.

Let us next apply the formalism developed above to discuss decays of (neutral) B mesons that are particularly important for the physics programme of the B factories.

6 BENCHMARK MODES FOR THE B FACTORIES

6.1 Exploring CP violation through $B \to J/\psi K$

6.1.1 Amplitude structure and CP asymmetries

One of the most prominent B decays is given by $B_d \to J/\psi K_{\rm S}$. If we take the CP parities of the J/ψ and $K_{\rm S}$ into account,[5] and note that these mesons are produced in a P wave with angular momentum $L = 1$, we find that the final state of this transition is an eigenstate of the CP operator, with eigenvalue

$$\underbrace{(+1)}_{J/\psi} \times \underbrace{(+1)}_{K_{\rm S}} \times \underbrace{(-1)^1}_{L=1} = -1.$$

As can be seen in Fig. 13, $B_d^0 \to J/\psi K_{\rm S}$ originates from $\bar{b} \to \bar{c}c\bar{s}$ quark-level decays, and receives contributions from tree and penguin topologies (see the classification in Subsection 4.3.1). Consequently, we may write the decay amplitude as follows [87]:

$$A(B_d^0 \to J/\psi K_{\rm S}) = \lambda_c^{(s)} \left(A_{\rm T}^{c'} + A_{\rm P}^{c'}\right) + \lambda_u^{(s)} A_{\rm P}^{u'} + \lambda_t^{(s)} A_{\rm P}^{t'}, \tag{141}$$

where $A_{\rm T}^{c'}$ corresponds to the tree process in Fig. 13, and the strong amplitudes $A_{\rm P}^{q'}$ describe the penguin topologies with internal q-quark exchanges ($q \in \{u, c, t\}$), including QCD and EW penguins; the primes remind us that we are dealing with a $\bar{b} \to \bar{s}$ transition. Finally, the

$$\lambda_q^{(s)} \equiv V_{qs} V_{qb}^* \tag{142}$$

are CKM factors. If we eliminate now $\lambda_t^{(s)}$ through (83) and apply the Wolfenstein parametrization, we straightforwardly arrive at

$$A(B_d^0 \to J/\psi K_{\rm S}) \propto \left[1 + \lambda^2 a e^{i\theta} e^{i\gamma}\right], \tag{143}$$

[5] Here we neglect the tiny indirect CP violation in the neutral kaon system.

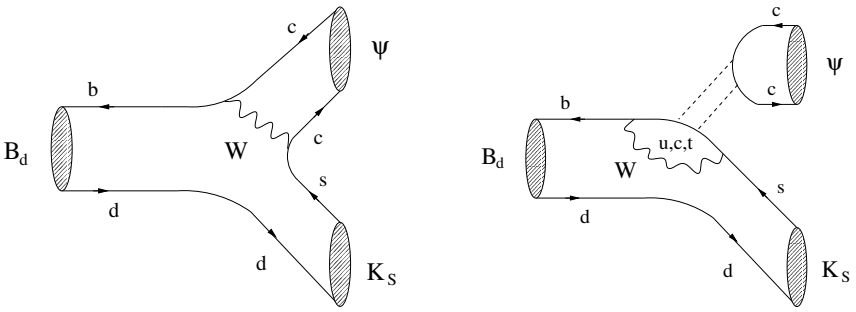


Fig. 13: Feynman diagrams contributing to $B_d^0 \to J/\psi K_{\rm S}$. The dashed lines in the penguin topology represent a colour-singlet exchange.

where

$$ae^{i\vartheta} \equiv \left(\frac{R_b}{1-\lambda^2}\right)\left[\frac{A_{\rm P}^{u'} - A_{\rm P}^{t'}}{A_{\rm T}^{c'} + A_{\rm P}^{c'} - A_{\rm P}^{t'}}\right] \tag{144}$$

is a hadronic parameter that is a measure for the ratio of the $B_d^0 \to J/\psi K_{\rm S}$ penguin to tree contributions. Using the results derived in Subsection 5.2, we obtain

$$\xi_{\psi K_{\rm S}}^{(d)} = +e^{-i\phi_d}\left[\frac{1+\lambda^2 a e^{i\vartheta}e^{-i\gamma}}{1+\lambda^2 a e^{i\vartheta}e^{+i\gamma}}\right]. \tag{145}$$

Unfortunately, the parameter $ae^{i\vartheta}$ can only be estimated with large hadronic uncertainties. However, since it enters (145) in a doubly Cabibbo-suppressed way, its impact on the CP-violating observables is practically negligible. We can put this statement on a more quantitative basis by making the plausible assumption that $a = \mathcal{O}(\bar\lambda) = \mathcal{O}(0.2) = \mathcal{O}(\lambda)$, where $\bar\lambda$ is a 'generic' expansion parameter. Applying now (127) yields

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to J/\psi K_{\rm S}) = 0 + \mathcal{O}(\overline{\lambda}^3) \tag{146}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S}) = -\sin\phi_d + \mathcal{O}(\overline{\lambda}^3) \stackrel{\rm SM}{=} -\sin 2\beta + \mathcal{O}(\overline{\lambda}^3). \tag{147}$$

These expressions are one of the most important applications of the general features that we discussed in the context of (138)–(140).

6.1.2 Experimental status and theoretical uncertainties

Looking at (147), we observe that the mixing-induced CP violation in $B_d \to J/\psi K_{\rm S}$ allows us to determine $\sin 2\beta$ in an essentially *clean* manner [88]. Because of this feature, this transition is referred to as the 'golden' mode to measure the angle β of the UT. After important first steps by the OPAL, CDF and ALEPH collaborations, the $B_d \to J/\psi K_{\rm S}$ mode (and similar decays) eventually led, in 2001, to the observation of CP violation in the B system [4], [5]. The current status of $\sin 2\beta$ is given as follows:

$$\sin 2\beta = \begin{cases} 0.741 \pm 0.067 \pm 0.033 & \text{(BaBar [89])} \\ 0.733 \pm 0.057 \pm 0.028 & \text{(Belle [90]),} \end{cases} \tag{148}$$

yielding the world average

$$\sin 2\beta = 0.736 \pm 0.049. \tag{149}$$

On the other hand, the CKM fits of the UT described in Subsection 2.7 imply the ranges in (43), where the one for β can be converted into

$$0.6 \lesssim \sin 2\beta \lesssim 0.9, \tag{150}$$

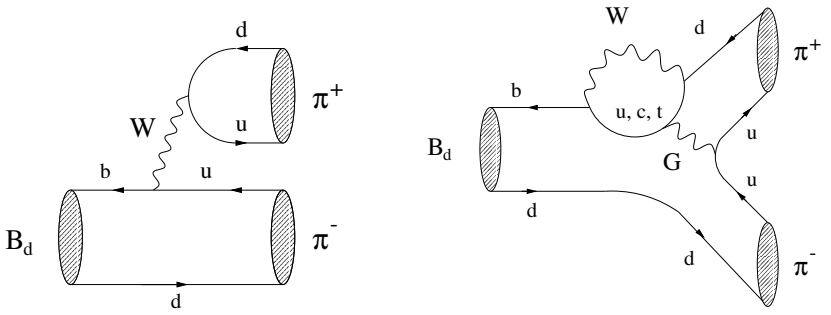


Fig. 14: Feynman diagrams contributing to $B_d^0 \to \pi^+\pi^-$.

which agrees well with the *direct* determination summarized in (149).

As far as the theoretical accuracy of (146) and (147) is concerned, the corrections, which originate from the penguin contributions and are at most of $\mathcal{O}(1\%)$,[6] are not yet an issue. However, in the era of the LHC [19], the experimental accuracy will be so tremendous that we have to start to deal with these terms. A possibility to control them is provided by the $B_s \to J/\psi K_{\rm S}$ channel, which can be combined with $B_d \to J/\psi K_{\rm S}$ through flavour-symmetry relations [87]. Moreover, also the direct CP violation in the $B \to J/\psi K$ system allows us to probe such penguin effects [71], where a combined analysis of the neutral $B_d \to J/\psi K_{\rm S}$ and charged $B^\pm \to J/\psi K^\pm$ modes provides the whole picture [91]; the current B-factory data for the corresponding direct CP asymmetries are consistent with zero. In a very recent analysis [92], this issue was also addressed from a more theoretical point of view. The corresponding estimates lead to tiny corrections at the 10^{-3} level, in accordance with the picture developed in Ref. [91].

Although the agreement between (149) and the results of the CKM fits is striking, it should not be forgotten that NP may—in principle—nevertheless hide in $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S})$. The point is that the key quantity is actually ϕ_d, which is fixed through $\sin\phi_d = 0.736 \pm 0.049$ up to a twofold ambiguity,

$$\phi_d = (47 \pm 4)^\circ \ \vee \ (133 \pm 4)^\circ. \tag{151}$$

Here the former solution would be in perfect agreement with CKM fits, implying $40^\circ \lesssim 2\beta \stackrel{\rm SM}{=} \phi_d \lesssim 60^\circ$, whereas the latter would correspond to NP. The two solutions can be distinguished through a measurement of the sign of $\cos\phi_d$: in the case of $\cos\phi_d = +0.7 > 0$, we would conclude $\phi_d = 47^\circ$, whereas $\cos\phi_d = -0.7 < 0$ would point towards $\phi_d = 133^\circ$, i.e. to NP. There are several strategies on the market to resolve the twofold ambiguity in the extraction of ϕ_d [93]. Unfortunately, they are rather challenging from a practical point of view. For instance, in the $B \to J/\psi K$ system, $\cos\phi_d$ can be extracted from the time-dependent angular distribution of the decay products of $B_d \to J/\psi[\to \ell^+\ell^-]K^*[\to \pi^0 K_{\rm S}]$, if the sign of a hadronic parameter $\cos\delta$ involving a strong phase δ is fixed through factorization [94], [95].

6.2 Exploring CP violation through $B \to \pi\pi$

6.2.1 Amplitude structure and CP asymmetries

Another benchmark mode for the B factories is the decay $B_d^0 \to \pi^+\pi^-$, which is a transition into a CP eigenstate with eigenvalue $+1$, and originates from $\bar{b} \to \bar{u}u\bar{d}$ quark-level processes, as can be seen in Fig. 14. In analogy to (141), the decay amplitude can be written in the following form [96]:

$$A(B_d^0 \to \pi^+\pi^-) = \lambda_u^{(d)}\,(A_{\rm T}^u + A_{\rm P}^u) + \lambda_c^{(d)} A_{\rm P}^c + \lambda_t^{(d)} A_{\rm P}^t. \tag{152}$$

If we use again (83) to eliminate the CKM factor $\lambda_t^{(d)} = V_{td}V_{tb}^*$ and apply once more the Wolfenstein parametrization, we obtain

$$A(B_d^0 \to \pi^+\pi^-) \propto \left[e^{i\gamma} - de^{i\theta}\right], \tag{153}$$

[6] In this case, the penguin topologies would *not* be suppressed with respect to the tree contributions, i.e. $a = \mathcal{O}(1)$.

where the hadronic parameter

$$de^{i\theta} \equiv \frac{1}{R_b}\left[\frac{A_{\rm P}^c - A_{\rm P}^t}{A_{\rm T}^u + A_{\rm P}^u - A_{\rm P}^t}\right] \tag{154}$$

is a measure for the ratio of the $B_d \to \pi^+\pi^-$ penguin to tree amplitudes. The formalism discussed in Subsection 5.2 then implies

$$\xi^{(d)}_{\pi^+\pi^-} = -e^{-i\phi_d}\left[\frac{e^{-i\gamma} - de^{i\theta}}{e^{+i\gamma} - de^{i\theta}}\right]. \tag{155}$$

In contrast to the expression for the $B^0_d \to J/\psi K_{\rm S}$ counterpart given in (145), the hadronic parameter $de^{i\theta}$, which suffers from large theoretical uncertainties, does *not* enter in (155) in a doubly Cabibbo-suppressed way. This feature is at the basis of the famous 'penguin problem' in $B_d \to \pi^+\pi^-$, which was addressed in many papers over the recent years (see, for instance, Refs. [97]– [102]). If we had negligible penguin contributions in this channel, i.e. $d = 0$, the corresponding CP-violating observables were simply given as follows:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = 0 \tag{156}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = \sin(\phi_d + 2\gamma) \stackrel{\rm SM}{=} \sin(\underbrace{2\beta + 2\gamma}_{2\pi - 2\alpha}) = -\sin 2\alpha. \tag{157}$$

Consequently, $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-)$ would allow us to determine α. However, in the general case of $d \neq 0$, we obtain formulae with the help of (127) and (155), which are considerably more complicated:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = -\left[\frac{2d\sin\theta\sin\gamma}{1 - 2d\cos\theta\cos\gamma + d^2}\right] \tag{158}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = \frac{\sin(\phi_d + 2\gamma) - 2d\cos\theta\sin(\phi_d + \gamma) + d^2\sin\phi_d}{1 - 2d\cos\theta\cos\gamma + d^2}. \tag{159}$$

We observe that actually the phases ϕ_d and γ enter directly in the $B_d \to \pi^+\pi^-$ observables, and not α. Consequently, since ϕ_d can be fixed straightforwardly through the mixing-induced CP violation in the "golden" mode $B_d \to J/\psi K_{\rm S}$, as we have seen in (147), we may use $B_d \to \pi^+\pi^-$ to probe γ. This is advantageous to deal with penguins and possible NP effects.

6.2.2 Experimental status and the 'B → ππ puzzle'

Measurements of the $B_d \to \pi^+\pi^-$ CP asymmetries are already available:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = \begin{cases} -0.19 \pm 0.19 \pm 0.05 & \text{(BaBar [103])} \\ -0.77 \pm 0.27 \pm 0.08 & \text{(Belle [104])} \end{cases} \tag{160}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = \begin{cases} +0.40 \pm 0.22 \pm 0.03 & \text{(BaBar [103])} \\ +1.23 \pm 0.41^{+0.07}_{-0.08} & \text{(Belle [104]).} \end{cases} \tag{161}$$

Unfortunately, the BaBar and Belle results are not fully consistent with each other, although both experiments point towards the same signs, and the last BaBar update of $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-)$ has moved towards Belle. In Ref. [86], the Heavy Flavour Averaging Group (HFAG) gave the following averages:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = -0.38 \pm 0.16 \tag{162}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = +0.58 \pm 0.20. \tag{163}$$

Direct CP violation at this level would require large penguin contributions with large CP-conserving strong phases, as is evident from (158). As we will see in Subsection 8.3.3, the CP asymmetries in (162) and (163) can be converted into the angle γ of the UT, with a result around 65°, in remarkable accordance with the SM picture [48], [105].

In addition to the decays $B_d \to \pi^+\pi^-$ and $B^\pm \to \pi^\pm\pi^0$, the B factories have recently reported the observation of the $B_d \to \pi^0\pi^0$ channel, with the following CP-averaged branching ratios:

$$\mathrm{BR}(B_d \to \pi^0\pi^0) = \begin{cases} (2.1 \pm 0.6 \pm 0.3) \times 10^{-6} & \text{(BaBar [106])} \\ (1.7 \pm 0.6 \pm 0.2) \times 10^{-6} & \text{(Belle [107]);} \end{cases} \tag{164}$$

CP-averaged branching ratios of this kind are generally defined through

$$\mathrm{BR} \equiv \frac{1}{2}\left[\mathrm{BR}(B \to f) + \mathrm{BR}(\bar{B} \to \bar{f})\right]. \tag{165}$$

These measurements represent quite a challenge for theory. For example, in a recent state-of-the-art calculation within QCD factorization [108], a $B_d \to \pi^0\pi^0$ branching ratio that is about six times smaller is favoured, whereas the calculation of $B_d \to \pi^+\pi^-$ points towards a branching ratio about two times larger than the current experimental average. On the other hand, the calculation of $B^\pm \to \pi^\pm\pi^0$ reproduces the data rather well. This '$B \to \pi\pi$ puzzle' is reflected by the following quantities [48], [54]:

$$R_{+-}^{\pi\pi} \equiv 2\left[\frac{\mathrm{BR}(B^\pm \to \pi^\pm\pi^0)}{\mathrm{BR}(B_d \to \pi^+\pi^-)}\right]\frac{\tau_{B_d^0}}{\tau_{B^+}} = 2.12 \pm 0.37 \tag{166}$$

$$R_{00}^{\pi\pi} \equiv 2\left[\frac{\mathrm{BR}(B_d \to \pi^0\pi^0)}{\mathrm{BR}(B_d \to \pi^+\pi^-)}\right] = 0.83 \pm 0.23; \tag{167}$$

the central values calculated within QCD factorization give $R_{+-}^{\pi\pi} = 1.24$ and $R_{00}^{\pi\pi} = 0.07$ [108]. As was discussed in detail in Refs. [48], [54], the $B \to \pi\pi$ puzzle can straightforwardly be accommodated within the SM through non-factorizable hadronic interference effects.[7] If we use

$$\phi_d = (47 \pm 4)^\circ, \quad \gamma = (65 \pm 7)^\circ, \tag{168}$$

as in the SM [41], this analysis allows us to convert the $B \to \pi\pi$ data into certain hadronic parameters. In particular, we obtain

$$d = 0.48^{+0.35}_{-0.22}, \quad \theta = +(138^{+19}_{-23})^\circ, \tag{169}$$

whereas QCD factorization favours $d \sim 0.3$ and $\theta \sim 180^\circ$. Moreover, the CP-violating observables of $B_d \to \pi^0\pi^0$ can be predicted, with the result

$$\mathcal{A}_{\mathrm{CP}}^{\mathrm{dir}}(B_d \to \pi^0\pi^0) = -0.41^{+0.35}_{-0.17}, \quad \mathcal{A}_{\mathrm{CP}}^{\mathrm{mix}}(B_d \to \pi^0\pi^0) = -0.55^{+0.43}_{-0.45}. \tag{170}$$

We shall return to $B_d \to \pi^+\pi^-$ in Subsection 8.3, in the context of $B_s \to K^+K^-$ [96].

6.3 Exploring CP violation through $B \to \phi K$

6.3.1 Amplitude structure and CP asymmetries

Another important mode for the testing of the KM mechanism of CP violation is provided by $B_d \to \phi K_{\mathrm{S}}$, which is—in analogy to $B_d \to J/\psi K_{\mathrm{S}}$—a decay into a CP-odd final state. As can be seen in Fig. 15, $B_d^0 \to \phi K_{\mathrm{S}}$ originates from $\bar{b} \to \bar{s}s\bar{s}$ quark-level processes, i.e. is a pure penguin mode. Consequently, $B_d^0 \to \phi K_{\mathrm{S}}$ and its charged counterpart $B^+ \to \phi K^+$ are governed by QCD penguin topologies [112], but also EW penguins have a sizeable impact because of the large top-quark mass [70], [113]. Using the same notation as above, we may write the $B_d^0 \to \phi K_{\mathrm{S}}$ decay amplitude within the SM as follows:

$$A(B_d^0 \to \phi K_{\mathrm{S}}) = \lambda_u^{(s)}\tilde{A}_{\mathrm{P}}^{u\prime} + \lambda_c^{(s)}\tilde{A}_{\mathrm{P}}^{c\prime} + \lambda_t^{(s)}\tilde{A}_{\mathrm{P}}^{t\prime}. \tag{171}$$

[7] Similar conclusions were also drawn very recently in Refs. [109], [110]. In Ref. [109], also the phenomenological implications of bounds on the UT that can be derived from the CP-violating $B_d \to \pi^+\pi^-$ observables, as pointed out in Ref. [111], were discussed.

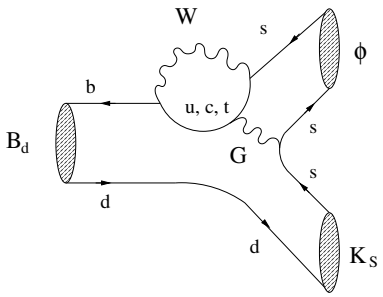


Fig. 15: Feynman diagrams contributing to $B_d \to \phi K_{\rm S}$.

Applying now once more (83) to eliminate the CKM factor $\lambda_t^{(s)}$, we obtain

$$A(B_d^0 \to \phi K_{\rm S}) \propto \left[1 + \lambda^2 b e^{i\Theta} e^{i\gamma}\right], \tag{172}$$

so that

$$\xi_{\phi K_{\rm S}}^{(d)} = +e^{-i\phi_d}\left[\frac{1 + \lambda^2 b e^{i\Theta} e^{-i\gamma}}{1 + \lambda^2 b e^{i\Theta} e^{+i\gamma}}\right], \tag{173}$$

with

$$b e^{i\Theta} = \left(\frac{R_b}{1-\lambda^2}\right)\left[\frac{\tilde{A}_{\rm P}^{u'} - \tilde{A}_{\rm P}^{t'}}{\tilde{A}_{\rm P}^{c'} - \tilde{A}_{\rm P}^{t'}}\right]. \tag{174}$$

The theoretical estimates of the hadronic parameter $b e^{i\Theta}$ suffer from large uncertainties. However, since this parameter enters (173) in a doubly Cabibbo-suppressed way, we obtain the simple expressions

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \phi K_{\rm S}) = 0 + \mathcal{O}(\lambda^2) \tag{175}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \phi K_{\rm S}) = -\sin\phi_d + \mathcal{O}(\lambda^2), \tag{176}$$

where we made the plausible assumption that $b = \mathcal{O}(1)$. On the other hand, the mixing-induced CP asymmetry of the 'golden' mode $B_d \to J/\psi K_{\rm S}$ measures also $-\sin\phi_d$ [see (147)]. Consequently, we arrive at the following relation [71], [114], [115], [116]:

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \phi K_{\rm S}) = \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S}) + \mathcal{O}(\lambda^2), \tag{177}$$

which offers a very interesting test of the SM description of CP violation. In order to obtain the whole picture and to search for NP systematically, it is useful to perform a combined analysis of the neutral $B_d \to \phi K_{\rm S}$ and the charged $B^\pm \to \phi K^\pm$ modes [116] (for a recent update, see Ref. [48]).

6.3.2 Experimental status

The experimental status of the CP-violating $B_d \to \phi K_{\rm S}$ observables is given as follows [117]:[8]

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \phi K_{\rm S}) = \begin{cases} +0.01 \pm 0.33 \pm 0.10 & \text{(BaBar [118])} \\ +0.15 \pm 0.29 \pm 0.07 & \text{(Belle [119])} \end{cases} \tag{178}$$

$$\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \phi K_{\rm S}) = \begin{cases} -0.47 \pm 0.34^{+0.06}_{-0.08} & \text{(BaBar [118])} \\ +0.96 \pm 0.50^{+0.11}_{-0.09} & \text{(Belle [119]),} \end{cases} \tag{179}$$

Since we have, on the other hand, $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S}) = -0.736 \pm 0.049$, we arrive at a puzzling situation, which has already stimulated many speculations about NP effects in the decay $B_d \to \phi K_{\rm S}$ (see, for instance, Ref. [120]). However, because of the very unsatisfactory current experimental picture, it seems too early to get too excited by the possibility of having a violation of the SM relation (177). It will be very interesting to observe how the B-factory data will evolve, and to keep also an eye on $B_d \to \eta' K_{\rm S}$ and other related modes.

[8] Note that the very recent BaBar update in Ref. [118] uses also $B_d \to \phi K_{\rm L}$ to extract the CP asymmetries of $B_d^0 \to \phi K^0$.

6.4 Manifestations of New Physics

6.4.1 New-Physics effects in B^0_d–$\bar{B}^0_d$ mixing

As we have seen in Subsection 5.1, B^0_d–$\bar{B}^0_d$ mixing originates in the SM from box diagrams, which are characterized by the Inami–Lim function $S_0(x_t)$. Concerning the impact of NP, it may enter B^0_d–$\bar{B}^0_d$ mixing through new-particle exchanges in the loop diagrams shown in Fig. 12, or through new FCNC processes arising at the tree level. The impact on the mixing parameters is twofold:

– The mass difference of the mass eigenstates is generalized as

$$\Delta M_d = \Delta M_d^{\rm SM} + \Delta M_d^{\rm NP}, \tag{180}$$

so that the NP contribution would affect the determination of the UT side R_t through (111).

– The CP-violating weak mixing phase is generalized as

$$\phi_d = \phi_d^{\rm SM} + \phi_d^{\rm NP} = 2\beta + \phi_d^{\rm NP}, \tag{181}$$

so that NP may enter the mixing-induced CP asymmetries through $\phi_d^{\rm NP}$.

On the basis of dimensional arguments borrowed from effective field theory (see, for instance, Refs. [47], [91]), and in specific NP scenarios, the following pattern may—in principle—be possible:

$$\Delta M_d^{\rm NP}/\Delta M_d^{\rm SM} \sim 1, \quad \phi_d^{\rm NP}/\phi_d^{\rm SM} \sim 1. \tag{182}$$

The same is true for the case of B^0_s–$\bar{B}^0_s$ mixing, which may be significantly affected by NP as well.[9]

6.4.2 New-Physics effects in decay amplitudes

Another way for NP to manifest itself is through contributions to decay amplitudes. If the decay does *not* arise at the tree level in the SM, we may have potentially large NP effects. In particular, NP may enter through new particles running in the loops, or through new FCNC processes arising at the tree level. An important example for such decays is given by the $B \to \phi K$ system, which is governed by $\bar{b} \to \bar{s}s\bar{s}$ penguin processes, as we have seen above. On the basis of general dimensional arguments [116], and in specific NP scenarios [120], significant effects may in fact arise in the $B \to \phi K$ amplitudes. The B-factory data may already indicate the presence of such a kind of NP, although it is too early to draw definite conclusions on this exciting possibility.

On the other hand, if a transition is dominated by a SM tree contribution, the impact of NP on the decay amplitude is generally small. An important example of this feature is given by the decay $B^0_d \to J/\psi K_{\rm S}$, which is governed by the $\bar{b} \to \bar{c}c\bar{s}$ process, arising at the tree level in the SM. Generic dimensional arguments then indicate that we may have NP effects at the $B \to J/\psi K$ amplitude level of at most $\mathcal{O}(10\%)$ for a NP scale in the TeV regime. In order to search systematically for such effects, it is useful to perform a combined analysis of the neutral and charged $B \to J/\psi K$ modes, and to introduce appropriate observable combinations [91]; the current B-factory data do not indicate any anomaly (for a recent update, see Ref. [48]). Since the determination of ϕ_d from the mixing-induced CP violation in $B_d \to J/\psi K_{\rm S}$ is very robust under NP, we may use the corresponding experimental result as an input for other studies of CP violation, as we have noted above.

6.4.3 Back to the status of the B^0_d–$\bar{B}^0_d$ mixing phase ϕ_d

Let us now briefly come back to the two solutions for ϕ_d in (151). In this context, it is interesting to note that an upper bound on ϕ_d is implied by an upper bound on $R_b \propto |V_{ub}/V_{cb}|$, as can straightforwardly be seen in Fig. 4. To be specific, we have

$$\sin\beta_{\rm max} = R_b^{\rm max}, \tag{183}$$

[9] Let us note that also D^0–$\bar{D}^0$ mixing offers an interesting probe to search for NP. Within the SM, this phenomenon is tiny, but it may be enhanced by the presence of NP. A similar comment applies to the CP-violating effects in D-meson decays. For a recent overview, we refer the reader to Ref. [121], and the references therein.

which yields $(\phi_d)^{\rm SM}_{\rm max} \sim 57^\circ$ for $R_b^{\rm max} \sim 0.48$. Since the determination of R_b from the semileptonic (tree-level) decays discussed in Subsection 4.2 is not expected to be sensitive to NP, $\phi_d \sim 133^\circ$ would require CP-violating NP contributions to B^0_d–$\bar{B}^0_d$ mixing. An interesting connection between the two solutions for ϕ_d and the UT angle γ is provided by the CP asymmetries of $B_d \to \pi^+\pi^-$ [47, 105]. We shall return to this feature in Section 8.

6.4.4 Models with minimal flavour violation

An interesting scenario for NP is provided by the simplest class of extensions of the SM. It is represented by models with 'minimal flavour violation' (MFV), which we may characterize as follows [122], [123] (for alternative definitions, see Refs. [124], [125]):

- All flavour-changing transitions are still governed by the CKM matrix, in particular no new phases.
- The only relevant operators are those already present in the SM.

Important examples are the Two-Higgs-Doublet Model II, the constrained MSSM (if $\tan\overline{\beta} = v_2/v_1$ is not too large), and models with universal extra dimensions [122]. As was pointed out in Ref. [123], a 'universal unitarity triangle' can be constructed for such MFV models with the help of those quantities that are not affected by the corresponding NP contributions. Following these lines, the 'true' values of $\bar{\rho}$ and $\bar{\eta}$ can still be determined in a transparent manner, despite the presence of NP.

Because of the items listed above, all SM expressions for decay amplitudes, as well as for particle–antiparticle mixing, can be generalized to the MFV models through a straightforward replacement of the initial Wilson coefficients for the renormalization-group evolution from $\mu = \mathcal{O}(M_W)$ down to appropriate 'low-energy' scales μ through characteristic NP coefficients. If we consider, for example, B^0_d–$\bar{B}^0_d$ mixing, we just have to make the following substitution for the Inami–Lim function $S_0(x_t)$:

$$S_0(x_t) \to S(v), \tag{184}$$

where v, which equals $x_t = m_t^2/M_W^2$ in the SM, denotes collectively the parameters of a given MFV model. Note that the *same* short-distance function governs also B^0_s–$\bar{B}^0_s$ mixing, as well as K^0–$\bar{K}^0$ mixing, so that it also enters the expression for the CP-violating observable ε.

Since no new phases appear in MFV models, one may think that the B^0_d–$\bar{B}^0_d$ mixing phase introduced in (136) would not be affected in such scenarios. However, because of a subtlety, this is actually not the case [126]. If we look at (103), we observe that the sign of $S_0(x_t)$ enters implicitly ϕ_d; in (122) and (136), we have actually used the fact that $S_0(x_t)$ is *positive*. However, since $S_0(x_t)$ is now replaced by $S(v)$, which needs no longer be positive, the expression for ϕ_d in (136) is generalized as follows:

$$\phi_d = 2\beta + \arg(S(v)), \tag{185}$$

so that $\phi_d^{\rm NP}$ in (181) is either 0° or 180° for $S(v) > 0$ or $S(v) < 0$, respectively. Consequently, in the most general MFV case, the mixing-induced CP asymmetry of $B_d \to J/\psi K_{\rm S}$ is given by

$$-\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S}) \equiv a_{\psi K_{\rm S}} = {\rm sgn}(S(v)) \sin 2\beta. \tag{186}$$

On the other hand, $\Delta M_d^{\rm NP}$ in (180) may have a significant impact on ΔM_d. Similarly, also ε may be affected. However, since the NP effects enter ΔM_d and ε through the same generalized Inami–Lim function $S(v)$, we obtain correlations between these observables. In fact, the interplay between B^0_d–$\bar{B}^0_d$ mixing and ε in the CKM fits implies bounds on $\sin 2\beta$ [127]. Using (186), we may cancel the sign ambiguity due to ${\rm sgn}(S(v))$, and obtain the following lower bounds for $a_{\psi K_{\rm S}}$:

$$(a_{\psi K_{\rm S}})_{\rm min} = \begin{cases} 0.42 & (S(v) > 0 \text{ [127]}) \\ 0.69 & (S(v) < 0 \text{ [126]}). \end{cases} \tag{187}$$

Although these bounds were very exciting immediately after the first B-factory data for $a_{\psi K_{\rm S}}$ were announced, which favoured rather small values, they are now not effective because of the world average given in (149). We shall come back to NP scenarios with MFV in Subsections 9.1–9.3. For a very comprehensive discussion, we refer the reader to Ref. [122].

7 AMPLITUDE RELATIONS

As we have noted in Subsection 4.4, amplitude relations offer another important tool to explore CP violation. Let us now have a closer look at the corresponding strategies, where we distinguish between the use of theoretically clean and flavour-symmetry relations.

7.1 Theoretically clean relations

7.1.1 $B^\pm \to K^\pm D$

The prototype of the strategies using theoretically clean amplitude relations is provided by $B^\pm \to K^\pm D$ decays [128]. Looking at Fig. 16, we observe that $B^+ \to K^+\bar{D}^0$ and $B^+ \to K^+D^0$ are pure 'tree' decays. If we consider, in addition, the transition $B^+ \to D_+^0 K^+$, where D_+^0 denotes the CP eigenstate of the neutral D-meson system with eigenvalue $+1$,

$$|D_+^0\rangle = \frac{1}{\sqrt{2}}\left[|D^0\rangle + |\bar{D}^0\rangle\right], \tag{188}$$

we obtain interference effects, which are described by

$$\sqrt{2}A(B^+ \to K^+ D_+^0) = A(B^+ \to K^+D^0) + A(B^+ \to K^+\bar{D}^0) \tag{189}$$

$$\sqrt{2}A(B^- \to K^- D_+^0) = A(B^- \to K^-\bar{D}^0) + A(B^- \to K^-D^0). \tag{190}$$

These relations can be represented as two triangles in the complex plane. Since we have only to deal with tree-diagram-like topologies, we have moreover

$$A(B^+ \to K^+\bar{D}^0) = A(B^- \to K^-D^0) \tag{191}$$

$$A(B^+ \to K^+D^0) = A(B^- \to K^-\bar{D}^0) \times e^{2i\gamma}, \tag{192}$$

allowing a *theoretically clean* extraction of γ, as shown in Fig. 17. Unfortunately, these triangles are very squashed, since $B^+ \to K^+D^0$ is colour-suppressed with respect to $B^+ \to K^+\bar{D}^0$:

$$\left|\frac{A(B^+ \to K^+D^0)}{A(B^+ \to K^+\bar{D}^0)}\right| = \left|\frac{A(B^- \to K^-\bar{D}^0)}{A(B^- \to K^-D^0)}\right| \approx \frac{1}{\lambda}\frac{|V_{ub}|}{|V_{cb}|} \times \frac{a_2}{a_1} \approx 0.4 \times 0.3 = \mathcal{O}(0.1), \tag{193}$$

where the phenomenological 'colour' factors were introduced in Subsection 4.3.3.

Another—more subtle—problem is related to the measurement of $\mathrm{BR}(B^+ \to K^+D^0)$. From the theoretical point of view, $D^0 \to K^-\ell^+\nu$ would be ideal to measure this tiny branching ratio. However, because of the huge background from semileptonic B decays, we must rely on Cabibbo-allowed hadronic $D^0 \to f_{\rm NE}$ decays, such as $f_{\rm NE} = \pi^+K^-$, ρ^+K^-, ..., i.e. have to measure

$$B^+ \to K^+D^0\,[\to f_{\rm NE}]. \tag{194}$$

Unfortunately, we then encounter another decay path into the *same* final-state $K^+f_{\rm NE}$ through

$$B^+ \to K^+\bar{D}^0\,[\to f_{\rm NE}], \tag{195}$$

where $\mathrm{BR}(B^+ \to K^+\bar{D}^0)$ is *larger* than $\mathrm{BR}(B^+ \to K^+D^0)$ by a factor of $\mathcal{O}(10^2)$, while $\bar{D}^0 \to f_{\rm NE}$ is doubly Cabibbo-suppressed, i.e. the corresponding branching ratio is suppressed with respect to the one of $D^0 \to f_{\rm NE}$ by a factor of $\mathcal{O}(10^{-2})$. Consequently, we obtain interference effects of $\mathcal{O}(1)$ between the decay chains in (194) and (195). If two different final states $f_{\rm NE}$ are considered, γ could—in principle—be extracted [129], although this determination would then be more involved than the original triangle approach presented in Ref. [128].

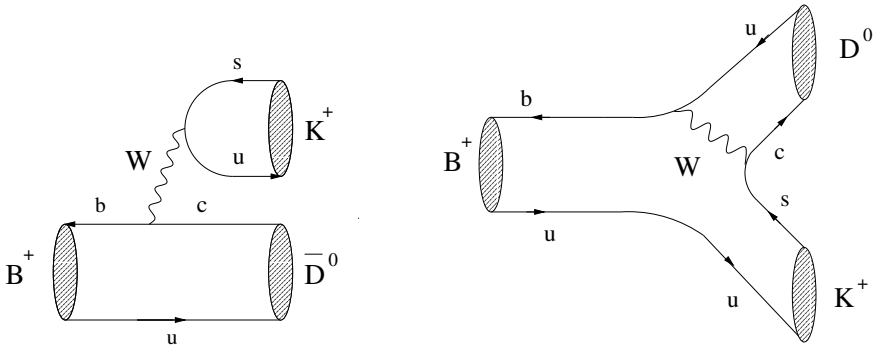


Fig. 16: Feynman diagrams contributing to $B^+ \to K^+\bar{D}^0$ and $B^+ \to K^+D^0$.

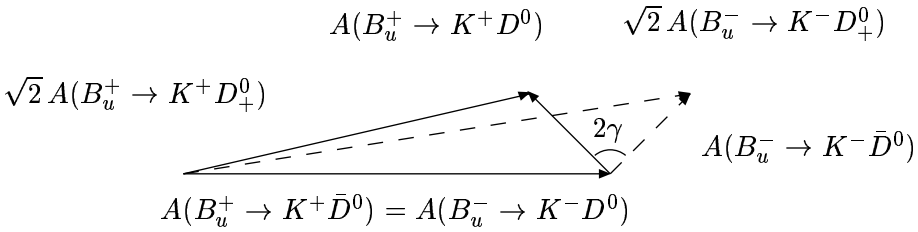


Fig. 17: The extraction of γ from $B^\pm \to K^\pm\{D^0, \bar{D}^0, D_+^0\}$ decays.

7.1.2 $B_c^\pm \to D_s^\pm D$

In addition to the 'conventional' $B_u^\pm$ mesons, there is yet another species of charged B mesons, the B_c-meson system, which consists of $B_c^+ \sim c\bar{b}$ and $B_c^- \sim b\bar{c}$. These mesons were observed by the CDF collaboration through their decay $B_c^+ \to J/\psi\ell^+\nu$, with the following mass and lifetime [130]:

$$M_{B_c} = (6.40 \pm 0.39 \pm 0.13)\,\text{GeV}, \quad \tau_{B_c} = (0.46^{+0.18}_{-0.16} \pm 0.03)\,\text{ps}. \tag{196}$$

Since a huge number of B_c mesons ($\sim 10^{10}$/year) will be produced at LHCb [19], the natural question arises of whether also the charged B_c-meson system provides a triangle approach to determine γ. Such a determination is actually offered by the decays $B_c^\pm \to D_s^\pm D$, which are the B_c-meson counterparts of the $B_u^\pm \to K^\pm D$ modes (see Fig. 18), and satisfy the following amplitude relations [131]:

$$\sqrt{2}A(B_c^+ \to D_s^+ D_+^0) = A(B_c^+ \to D_s^+ D^0) + A(B_c^+ \to D_s^+ \bar{D}^0) \tag{197}$$
$$\sqrt{2}A(B_c^- \to D_s^- D_+^0) = A(B_c^- \to D_s^- \bar{D}^0) + A(B_c^- \to D_s^- D^0), \tag{198}$$

with

$$A(B_c^+ \to D_s^+ \bar{D}^0) = A(B_c^- \to D_s^- D^0) \tag{199}$$
$$A(B_c^+ \to D_s^+ D^0) = A(B_c^- \to D_s^- \bar{D}^0) \times e^{2i\gamma}. \tag{200}$$

At first sight, everything is completely analogous to the $B_u^\pm \to K^\pm D$ case. However, there is an important difference [132], which becomes obvious by comparing the Feynman diagrams shown in Figs. 16 and 18: in the $B_c^\pm \to D_s^\pm D$ system, the amplitude with the rather small CKM matrix element V_{ub} is not colour-suppressed, while the larger element V_{cb} comes with a colour-suppression factor. Therefore, we obtain

$$\left|\frac{A(B_c^+ \to D_s^+ D^0)}{A(B_c^+ \to D_s^+ \bar{D}^0)}\right| = \left|\frac{A(B_c^- \to D_s^- \bar{D}^0)}{A(B_c^- \to D_s^- D^0)}\right| \approx \frac{1}{\lambda}\frac{|V_{ub}|}{|V_{cb}|} \times \frac{a_1}{a_2} \approx 0.4 \times 3 = \mathcal{O}(1), \tag{201}$$

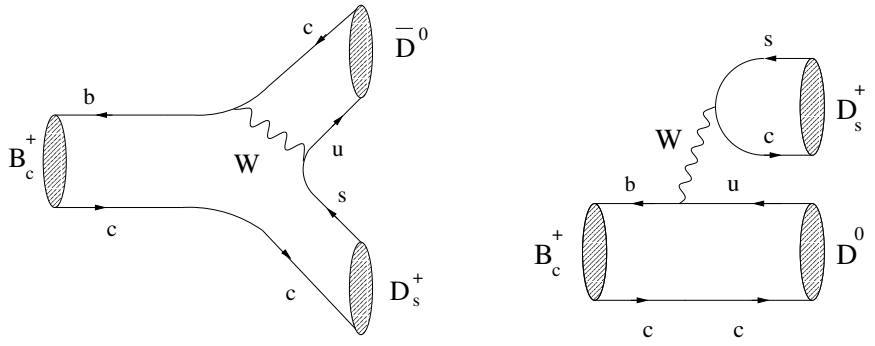


Fig. 18: Feynman diagrams contributing to $B_c^+ \to D_s^+ \bar{D}^0$ and $B^+ \to D_s^+ D^0$.

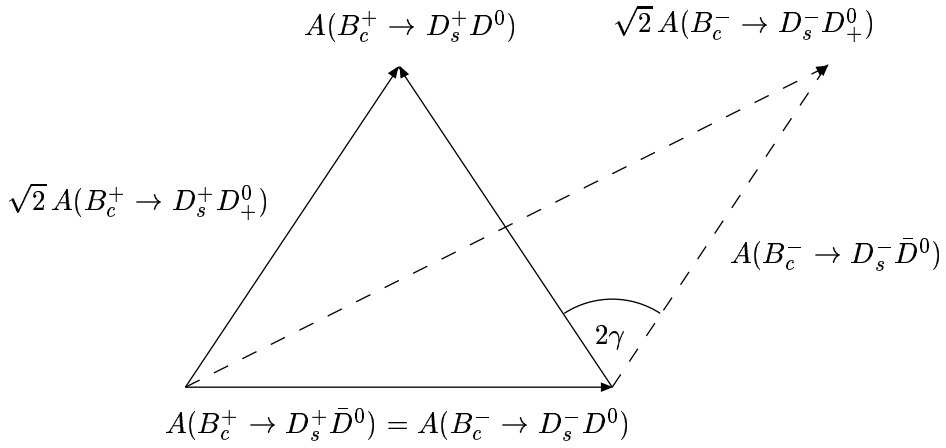


Fig. 19: The extraction of γ from $B_c^\pm \to D_s^\pm \{D^0, \bar{D}^0, D_+^0\}$ decays.

and conclude that the two amplitudes are similar in size. In contrast to this favourable situation, in the decays $B_u^\pm \to K^\pm D$, the matrix element V_{ub} comes with the colour-suppression factor, resulting in a very stretched triangle. The extraction of γ from the $B_c^\pm \to D_s^\pm D$ triangles is illustrated in Fig. 19, which should be compared with the squashed $B_u^\pm \to K^\pm D$ triangles shown in Fig. 17. Another important advantage is that the interference effects arising from $D^0, \bar{D}^0 \to \pi^+ K^-$ are practically unimportant for the measurement of $\text{BR}(B_c^+ \to D_s^+ D^0)$ and $\text{BR}(B_c^+ \to D_s^+ \bar{D}^0)$ since the B_c-decay amplitudes are of the same order of magnitude. Consequently, the $B_c^\pm \to D_s^\pm D$ decays provide—from the theoretical point of view—the ideal realization of the 'triangle' approach to determine γ. On the other hand, the practical implementation still appears to be challenging, although detailed experimental feasibility studies for LHCb are strongly encouraged. The corresponding branching ratios were recently estimated in Ref. [133], with a pattern in accordance with (201).

7.2 Flavour-symmetry relations: $B \to \pi K$

Let us now turn to amplitude relations that follow from the flavour symmetries of the strong interactions, which are—in contrast to the relations discussed in Subsection 7.1—not theoretically clean, but are nevertheless very useful to explore CP violation and to obtain insights into hadron dynamics. Here the prototype is provided by $B \to \pi K$ decays, which received a lot of attention in the B-physics community. Since a detailed discussion of the corresponding strategies is beyond the scope of these lectures, we address only their most important features and refer the interested reader to Ref. [26], where also a comprehensive list of references can be found.

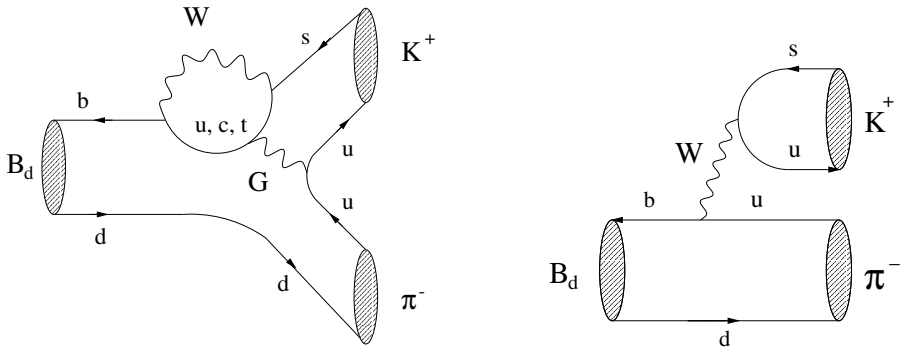


Fig. 20: Feynman diagrams contributing to $B^0_d \to \pi^- K^+$.

7.2.1 General features

In order to get more familiar with the $B \to \pi K$ modes, let us consider the decay $B^0_d \to \pi^- K^+$. As can be seen in Fig. 20, this channel receives contributions from penguin and tree topologies. Consequently, $B^0_d \to \pi^- K^+$ exhibits interference effects between the penguin and tree amplitudes, where the latter brings the angle γ of the UT into the game. Because of the small ratio $|V_{us}V^*_{ub}/(V_{ts}V^*_{tb})| \approx 0.02$, the QCD penguin topologies play the dominant role in this decay, despite their loop suppression. The ratio of the tree to the penguin amplitudes is generically expected at the 20% level. Interestingly, all $B \to \pi K$ modes are governed by their QCD penguin contributions. Because of the large top-quark mass, we have also to care about EW penguins:

- In the case of $B^0_d \to \pi^- K^+$ and $B^+ \to \pi^+ K^0$, these topologies contribute only in colour-suppressed form and are hence expected to play a minor role, thereby leading to contributions to the decay amplitudes of $\mathcal{O}(1\%)$.
- On the other hand, EW penguins may also contribute to $B^+ \to \pi^0 K^+$ and $B^0_d \to \pi^0 K^0$ in colour-allowed form, and may here even compete with the tree-diagram-like topologies, thereby leading to contributions to the decay amplitudes of $\mathcal{O}(20\%)$.

It can be shown that the isospin flavour symmetry of strong interactions implies the relation

$$\begin{aligned}\sqrt{2}A(B^+ \to \pi^0 K^+) + A(B^+ \to \pi^+ K^0) &= \sqrt{2}A(B^0_d \to \pi^0 K^0) + A(B^0_d \to \pi^- K^+) \\ &= -\Big[\underbrace{|T + C|e^{i\delta_{T+C}}e^{i\gamma}}_{\text{tree topologies}} + \underbrace{(P_{\rm ew} + P^{\rm C}_{\rm ew})}_{\text{EW penguins}}\Big] \propto \left[e^{i\gamma} - q\right], \end{aligned} \tag{202}$$

where the T ($P_{\rm ew}$) and C ($P^{\rm C}_{\rm ew}$) denote the amplitudes of the colour-allowed and colour-suppressed tree (EW penguin) topologies, respectively, δ_{T+C} is a CP-conserving strong phase, and the factors of $\sqrt{2}$ originate from the wave functions of the neutral pions. Note that the QCD penguin contributions cancel in this expression. A relation with an analogous phase structure can also be derived for the $B^+ \to \pi^+ K^0$, $B^0_d \to \pi^- K^+$ system.

7.2.2 Extraction of γ and strong phases

The $B \to \pi K$ observables allow us to determine the angle γ of the UT. Because of the isospin relation in (202), we may separately consider the following decay combinations to this end:

- The 'mixed' system of the charged $B^\pm \to \pi^\pm K$ and neutral $B_d \to \pi^\mp K^\pm$ modes [134]– [137].
- The system of the charged $B^\pm \to \pi^\pm K$, $B^\pm \to \pi^0 K^\pm$ modes [138]– [140].
- The system of the neutral $B_d \to \pi^0 K$, $B_d \to \pi^\mp K^\pm$ modes [140], [141].

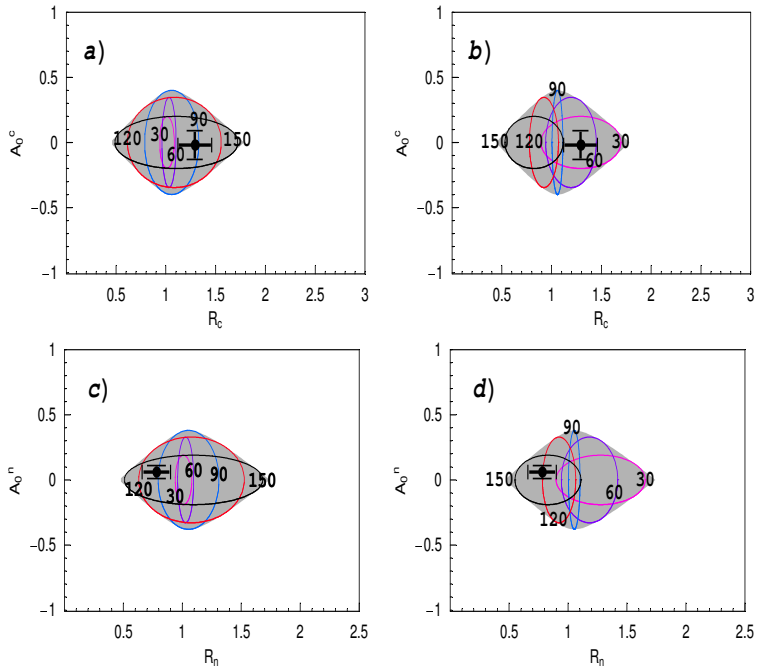


Fig. 21: The allowed regions in observable space of the charged [$r_{\rm c} = 0.20$; (a), (b)] and neutral [$r_{\rm n} = 0.19$; (c), (d)] $B \to \pi K$ systems for $q = 0.69$: in (a) and (c), we show also the contours for fixed values of γ, whereas we give the curves arising for fixed values of $|\delta_{\rm c}|$ and $|\delta_{\rm n}|$ in (b) and (d), respectively.

Correspondingly, we introduce the following sets of observables [140]:

$$\left\{ \begin{array}{c} R \\ A_0 \end{array} \right\} \equiv \left[\frac{{\rm BR}(B_d^0 \to \pi^- K^+) \pm {\rm BR}(\bar{B}_d^0 \to \pi^+ K^-)}{{\rm BR}(B^+ \to \pi^+ K^0) + {\rm BR}(B^- \to \pi^- \bar{K}^0)} \right] \frac{\tau_{B^+}}{\tau_{B_d^0}} \tag{203}$$

$$\left\{ \begin{array}{c} R_{\rm c} \\ A_0^{\rm c} \end{array} \right\} \equiv 2 \left[\frac{{\rm BR}(B^+ \to \pi^0 K^+) \pm {\rm BR}(B^- \to \pi^0 K^-)}{{\rm BR}(B^+ \to \pi^+ K^0) + {\rm BR}(B^- \to \pi^- \bar{K}^0)} \right] \tag{204}$$

$$\left\{ \begin{array}{c} R_{\rm n} \\ A_0^{\rm n} \end{array} \right\} \equiv \frac{1}{2} \left[\frac{{\rm BR}(B_d^0 \to \pi^- K^+) \pm {\rm BR}(\bar{B}_d^0 \to \pi^+ K^-)}{{\rm BR}(B_d^0 \to \pi^0 K^0) + {\rm BR}(\bar{B}_d^0 \to \pi^0 \bar{K}^0)} \right], \tag{205}$$

where the $R_{\rm (c,n)}$ and $A_0^{\rm (c,n)}$ refer to the plus and minus signs, respectively; the factors of 2 and $1/2$ are due to the wave functions of the neutral pions. In contrast to the observables in (203), those in (204) and (205) are significantly affected by EW penguins. We will return to this important feature below.

As noted in Ref. [140], all three $B \to \pi K$ systems can be described by the same set of formulae, just making straightforward replacements of variables. Let us first focus on the charged and neutral $B \to \pi K$ systems. For the parametrization of their observables, we employ the isospin relation mentioned above, and assume that certain rescattering effects are small; large rescattering processes would be indicated by large direct CP violation in $B^\pm \to \pi^\pm K$, which is not supported by the current B-factory average [86]:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B^\pm \to \pi^\pm K) = -0.02 \pm 0.06, \tag{206}$$

and by an enhancement of the $B \to KK$ branching ratios, which are already strongly constrained by the B-factory data as well (for detailed discussions, see Refs. [26], [48]). Following these lines, we may write

$$R_{\rm c,n} = {\rm function}(q, r_{\rm c,n}, \delta_{\rm c,n}, \gamma), \quad A_0^{\rm c,n} = {\rm function}(r_{\rm c,n}, \delta_{\rm c,n}, \gamma), \tag{207}$$

where the parameters q, $r_{\rm c,n}$ and $\delta_{\rm c,n}$ have the following meaning:

– q describes the ratio of the EW penguin to tree contributions (see (202)), which can be determined with the help of $SU(3)$ flavour-symmetry arguments, yielding the following SM result [48], [138]:

$$q|_{\rm SM} = 0.69 \times \left[\frac{0.086}{|V_{ub}/V_{cb}|}\right]. \tag{208}$$

– The parameters $r_{\rm c,n}$ measure the ratios of the tree to QCD penguin topologies, and can be fixed through $SU(3)$ arguments and the data for ${\rm BR}(B^\pm \to \pi^\pm\pi^0)$ [142], yielding $r_{\rm c,n} \sim 0.2$.

– The $\delta_{\rm c,n}$ are the CP-conserving strong phases between the tree and QCD penguin amplitudes.

Let us now consider either the charged or the neutral $B \to \pi K$ system. Since we may fix q and the corresponding $r_{\rm c,n}$ with the help of $SU(3)$ flavour-symmetry relations, the observables $R_{\rm c,n}$ and $A_0^{\rm c,n}$ depend only on the two 'unknown' parameters $\delta_{\rm c,n}$ and γ. If we vary them within their allowed ranges, i.e. $-180^\circ \le \delta_{\rm c,n} \le +180^\circ$ and $0^\circ \le \gamma \le 180^\circ$, we obtain an allowed region in the $R_{\rm c,n}$–$A_0^{\rm c,n}$ plane [105], [143]. Should the measured values of $R_{\rm c,n}$ and $A_0^{\rm c,n}$ fall outside this region, we would have an immediate signal for NP. On the other hand, should the measurements lie inside the allowed range, γ and $\delta_{\rm c,n}$ could be extracted. The value of γ thus obtained could then be compared with the results of other strategies, whereas the strong phase $\delta_{\rm c,n}$ would offer interesting insights into hadron dynamics.

In Fig. 21, we show the allowed regions in the $R_{\rm c,n}$–$A_0^{\rm c,n}$ planes following Ref. [105], where the crosses represent the averages of the B-factory data. As can be read off from the contours in these figures, both the charged and the neutral $B \to \pi K$ data favour $\gamma \gtrsim 90^\circ$, which would be in conflict with the results of the usual CKM fits, as summarized in (43). Moreover, we observe that the charged modes point towards $|\delta_{\rm c}| \lesssim 90^\circ$ (QCD factorization predicts $\delta_{\rm c}$ to be close to 0° [75], [108]), whereas the neutral decays prefer $|\delta_{\rm n}| \gtrsim 90^\circ$. Since we do not expect $\delta_{\rm c}$ to differ significantly from $\delta_{\rm n}$, we arrive at a 'puzzling' picture of the kind that was already pointed out in the year 2000 [141], and was recently reconsidered in Refs. [48], [54], [108], [144], [145], [146]. In the experimental values

$$R_{\rm c} = 1.17 \pm 0.12, \quad R_{\rm n} = 0.76 \pm 0.10, \tag{209}$$

this puzzle is reflected in particular by $R_{\rm n} < 1$, while $R_{\rm c} > 1$, as is now consistently favoured by the separate BaBar, Belle and CLEO data [86]. Concerning the mixed $B \to \pi K$ system, the data fall well into the SM region in observable space and do not indicate any 'anomalous' behaviour [105].

7.2.3 The '$B \to \pi K$ puzzle' and recent developments

Since $R_{\rm c}$ and $R_{\rm n}$ are affected significantly by colour-allowed EW penguins, whereas such topologies may only contribute to R in colour-suppressed form, the experimental pattern for these observables discussed above may be a manifestation of NP in the EW penguin sector [108], [141], [144]– [146], offering an attractive avenue for physics beyond the SM to enter the $B \to \pi K$ system [147], [148]. In order to deal with these effects quantitatively, we have to replace the parameter in (208), which characterizes the EW penguins in the SM, through a generalized parameter q, which may, in particular, also be associated with a CP-violating NP phase ϕ.

A detailed analysis of the $B \to \pi K$ puzzle was recently performed in Refs. [48], [54]. The starting point is the $B \to \pi\pi$ puzzle addressed in Subsection 6.2.2, which indicates that another hadronic parameter of the neutral $B \to \pi K$ system, $\rho_{\rm n} e^{i\theta_{\rm n}}$, is not as small as naïvely expected. However, using the $SU(3)$ flavour symmetry and plausible dynamical assumptions, it can be shown that we may fix all relevant hadronic $B \to \pi K$ parameters—including CP-conserving strong phases—through their $B \to \pi\pi$ counterparts, i.e. with the help of the B-factory data. Moreover, if we complement $B_d \to \pi^+\pi^-$ with $B_d \to \pi^\mp K^\pm$, we may also extract γ (see Subsection 8.3.3), with a result in excellent accordance with the range for γ in (168). Since EW penguins play a very minor role in $B \to \pi\pi$ and $B_d \to \pi^\mp K^\pm$ decays, these modes—and the parameters extracted from their observables—are essentially unaffected by NP in the EW penguin sector. Having all $B \to \pi K$ parameters at hand, we may then analyse the $B \to \pi K$ system in the SM.

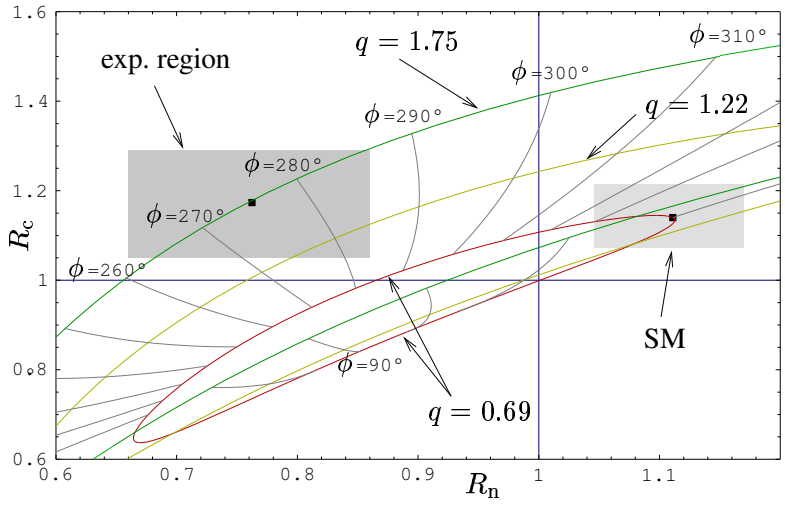


Fig. 22: The situation in the R_n–R_c plane, where the current experimental and SM ranges are indicated in grey. We show also contours for the EW penguin parameters $q = 0.69$, $q = 1.22$ and $q = 1.75$, with a NP phase $\phi \in [0°, 360°]$.

As far as the 'mixed' $B \to \pi K$ system is concerned, we obtain

$$R|_{\rm SM} = 0.943^{+0.033}_{-0.026}, \tag{210}$$

which agrees well with the experimental result $R = 0.91 \pm 0.07$ following from the averages compiled in Ref. [86]. Additional information is provided by direct CP violation. Whereas the direct CP asymmetry of $B^\pm \to \pi^\pm K$ vanishes within our working assumptions, in accordance with the experimental value in (206), we find

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^\mp K^\pm)\Big|_{\rm SM} = 0.140^{+0.139}_{-0.087}, \tag{211}$$

which is in agreement with the current B-factory average $\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^\mp K^\pm) = +0.095 \pm 0.028$.

In order to discuss the observables R_n and R_c, it is convenient to consider the R_n–R_c plane. Since all hadronic parameters are fixed through the $B \to \pi\pi$ data, these observables now depend only on the EW penguin parameters q and ϕ, where the SM is described by (208), corresponding to $\phi = 0°$. As can nicely be seen in Fig. 22, the pattern of the SM predictions

$$R_c|_{\rm SM} = 1.14^{+0.08}_{-0.07}, \quad R_n|_{\rm SM} = 1.11^{+0.06}_{-0.07} \tag{212}$$

is *not* in accordance with the current experimental picture (209), so that we are actually back at the $B \to \pi K$ puzzle described above. In this figure, we have also included various contours corresponding to different fixed values of q, where each point is parametrized through the value of $\phi \in [0°, 360°]$. We observe that we may in fact move to the experimental region for an enhanced value of $q \sim 1.8$ and $\phi \sim -90°$, where in particular the large CP-violating phase is in stark contrast to the SM. In order to put these observations on a more quantitative level, we may convert the experimental values of R_c and R_n into values of q and ϕ, with the following result:

$$q = 1.75^{+1.27}_{-0.99}, \quad \phi = -(85^{+11}_{-14})°. \tag{213}$$

Because of the large, non-vanishing value of ϕ, this scenario of NP would require new sources for CP violation, i.e. would *not* belong to the simple class of MFV models specified in Subsection 6.4.4. As far as direct CP violation in $B^\pm \to \pi^0 K^\pm$ is concerned, we obtain

$$\mathcal{A}_{\rm CP}^{\rm dir}(B^\pm \to \pi^0 K^\pm) = 0.04^{+0.37}_{-0.28} \tag{214}$$

in our NP scenario, in accordance with the experimental number $\mathcal{A}_{\rm CP}^{\rm dir}(B^\pm \to \pi^0 K^\pm) = 0.00 \pm 0.07$. As was pointed out in Ref. [134], also the CP asymmetries of $B_d \to \pi^0 K_{\rm S}$ are an important tool to explore the KM mechanism of CP violation, where the SM corresponds (for $\rho_{\rm n} = 0$) to the relations

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^0 K_{\rm S}) = 0, \quad \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^0 K_{\rm S}) = -\sin\phi_d = \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to J/\psi K_{\rm S}), \tag{215}$$

in analogy to (177). Recently, the BaBar collaboration reported the following results [149]:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^0 K_{\rm S}) = 0.40^{+0.27}_{-0.28} \pm 0.09, \quad \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^0 K_{\rm S}) = -0.48^{-0.38}_{+0.47} \pm 0.06. \tag{216}$$

Moreover, also a measurement of the direct CP asymmetry of the $B_d^0 \to \pi^0 K^0$ channel is available [86]:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d^0 \to \pi^0 K^0) = -0.03 \pm 0.36 \pm 0.09, \tag{217}$$

which is supposed to agree with the direct CP asymmetry in (216). Consequently, these experimental numbers are expected to change significantly in the future. On the other hand, the $B \to \pi\pi, \pi K$ analysis described above yields the predictions

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^0 K_{\rm S}) = +0.05^{+0.24}_{-0.29}, \quad \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^0 K_{\rm S}) = -0.99^{+0.04}_{-0.01}. \tag{218}$$

The measurement of these CP asymmetries will allow an interesting test of the NP scenario of enhanced EW penguins with a large CP-violating phase that is suggested by the $B \to \pi K$ puzzle. In this respect, it is important to consider also rare B and K decays, which offer particularly sensitive probes for the exploration of this kind of NP. We shall return to the corresponding NP effects in Subsection 9.4, where we shall also briefly address the impact on ${\rm Re}(\varepsilon'/\varepsilon)$, $B_d \to J/\psi K_{\rm S}$ and $B_d \to \phi K_{\rm S}$.

8 THE B_s-MESON SYSTEM

8.1 General features

8.1.1 Comparison of the B_d and B_s systems

At the e^+e^- B factories operating at the $\Upsilon(4S)$ resonance (BaBar and Belle), the B_s-meson system is not accessible since $\Upsilon(4S)$ states decay only into $B_{u,d}$ but not into B_s mesons.[10] On the other hand, plenty of B_s mesons will be produced at hadron colliders. Consequently, these particles are the 'El Dorado' for B-decay studies at run II of the Tevatron [18], and later on at the LHC [19]. There are important differences between the B_d and B_s systems:

- The B_s^0–$\bar{B}_s^0$ mixing phase is negligibly small in the SM,

$$\phi_s \equiv 2\arg(V_{ts}^* V_{tb}) = -2\delta\gamma = -2\lambda^2\eta = \mathcal{O}(-2^\circ), \tag{219}$$

  whereas $\phi_d \equiv 2\arg(V_{td}^* V_{tb}) = 2\beta = \mathcal{O}(50^\circ)$.
- A large mixing parameter x_s is expected in the SM,

$$x_s \equiv \frac{\Delta M_s}{\Gamma_s} = \mathcal{O}(20), \tag{220}$$

  whereas $x_d = 0.771 \pm 0.012$. Consequently, we have to deal with rapid B_s^0–$\bar{B}_s^0$ oscillations. The current experimental lower bound for the mass difference of the B_s mass eigenstates is given by $\Delta M_s > 14.5\,{\rm ps}^{-1}$, corresponding to $x_s > 20.8$ (95% C.L.) [85], [86].
- There may be a sizeable difference between the decay widths of the B_s mass eigenstates,

$$\frac{\Delta\Gamma_s}{\Gamma_s} = \mathcal{O}(-10\%), \tag{221}$$

  whereas $\Delta\Gamma_d/\Gamma_d$ is negligibly small, as we have seen in Subsection 5.1.2. The current CDF and LEP average is given by $\Delta\Gamma_s/\Gamma_s = -0.16^{+0.16}_{-0.15}$, $|\Delta\Gamma_s|/\Gamma_s < 0.54$ (95% C.L.) [85], [86].

[10] Operating these machines on the $\Upsilon(5S)$ resonance would also allow the production of B_s mesons.

8.1.2 Impact of ΔM_s on the unitarity triangle

As we discussed in Subsection 5.1, the mass differences of the B_q mass eigenstates satisfy

$$\Delta M_q \propto M_{B_q} \hat{B}_{B_q} f_{B_q}^2 |V_{tq}^* V_{tb}|^2. \tag{222}$$

In the B_d-meson case, this particular structure leads to (111), allowing us to determine the side R_t of the UT. To this end, in addition to the CKM parameter A (see (72)), also the non-perturbative quantity

$$\sqrt{\hat{B}_{B_d}} f_{B_d} = (235 \pm 33^{+0}_{-24})\,\text{MeV} \tag{223}$$

has to be known, where the numerical value follows from lattice QCD studies [41]; QCD sum rules give a similar picture [150]. On the other hand, if we apply the expressions for V_{cb} and V_{ts} in (21), we obtain

$$R_t \equiv \frac{1}{\lambda}\left|\frac{V_{td}}{V_{cb}}\right| = \frac{1}{\lambda}\left|\frac{V_{td}}{V_{ts}}\right|\left[1 + \mathcal{O}(\lambda^2)\right]. \tag{224}$$

Consequently, we may—up to corrections entering at the λ^2 level—determine R_t through the ratio $|V_{td}/V_{ts}|$. Using now (222) yields the following expression [82]:

$$R_t = 0.90\left[\frac{\xi}{1.24}\right]\sqrt{\frac{18.4\,\text{ps}^{-1}}{\Delta M_s}}\sqrt{\frac{\Delta M_d}{0.5\,\text{ps}^{-1}}}, \tag{225}$$

where

$$\xi \equiv \frac{\sqrt{\hat{B}_s} f_{B_s}}{\sqrt{\hat{B}_d} f_{B_d}} \tag{226}$$

is an $SU(3)$-breaking parameter; lattice QCD studies give

$$\xi = 1.18 \pm 0.04^{+0.12}_{-0}, \tag{227}$$

where $\xi = 1.24 \pm 0.08$ should be used for analyses of the UT, as discussed in [41]. In comparison with the quantity in (223) entering (111), the ratio in (226) is more favourable and represents an important aspect of current non-perturbative research [41]. Another advantage of (225) is that A, the Inami–Lim function $S_0(x_t)$, and the short-distance QCD correction factor η_B cancel in this expression. Interestingly, it allows us also to convert the lower experimental bound $\Delta M_s > 14.5\,\text{ps}^{-1}$ into the upper bound $R_t < 1.0 \times [\xi/1.24]$, which implies $\gamma \lesssim 90^\circ$, thereby excluding a large fraction of the $\bar{\rho}$–$\bar{\eta}$ plane.

8.1.3 $\Delta\Gamma_s$ and 'untagged' B_s rates

The width difference of the B_s-meson system may provide interesting studies of CP violation through 'untagged' B_s rates [151]– [153], which are defined as

$$\langle\Gamma(B_s(t) \to f)\rangle \equiv \Gamma(B_s^0(t) \to f) + \Gamma(\bar{B}_s^0(t) \to f), \tag{228}$$

and are characterized by the feature that we do not distinguish between initially, i.e. at time $t = 0$, present B_s^0 or $\bar{B}_s^0$ mesons. If we consider a final state f to which both a B_s^0 and a $\bar{B}_s^0$ may decay, and use the expressions in (123), we find

$$\langle\Gamma(B_s(t) \to f)\rangle \propto \left[\cosh(\Delta\Gamma_s t/2) - \mathcal{A}_{\Delta\Gamma}(B_s \to f)\sinh(\Delta\Gamma_s t/2)\right] e^{-\Gamma_s t}, \tag{229}$$

where $\mathcal{A}_{\Delta\Gamma}(B_s \to f) \propto \text{Re}\,\xi_f$ was introduced in (129). We observe that the rapidly oscillating $\Delta M_s t$ terms cancel, and that we may obtain information on the phase structure of the observable ξ_f, thereby providing valuable insights into CP violation. For instance, the untagged observables offered by the angular distribution of the $B_s \to K^{*+}K^{*-}, K^{*0}\bar{K}^{*0}$ decay products allow the determination of the UT angle γ, provided $\Delta\Gamma_s$ is actually sizeable [152]. Although B-decay experiments at hadron colliders should be able to resolve the B_s^0–$\bar{B}_s^0$ oscillations, untagged B_s rates are interesting in terms of efficiency, acceptance and purity.

8.2 $B_s \to J/\psi\phi$

This particularly promising channel is the B_s-meson counterpart of the 'golden' mode $B_d \to J/\psi K_{\rm S}$, as can be seen from the diagrams shown in Fig. 13, where we just have to replace the down spectator quark by a strange quark in order to obtain the $B_s \to J/\psi\phi$ diagrams. Consequently, this decay is described by a transition amplitude with a structure that is completely analogous to that of (143). On the other hand, in contrast to $B_d \to J/\psi K_{\rm S}$, the final state of $B_s \to J/\psi\phi$ is an admixture of different CP eigenstates, which can, however, be disentangled through an angular analysis of the $J/\psi[\to \ell^+\ell^-]\phi[\to K^+K^-]$ decay products [154], [155]. Their angular distribution exhibits tiny direct CP violation, whereas mixing-induced CP-violating effects allow the extraction of

$$\sin\phi_s + \mathcal{O}(\overline{\lambda}^3) = \sin\phi_s + \mathcal{O}(10^{-3}). \tag{230}$$

Since we have $\phi_s = -2\lambda^2\eta = \mathcal{O}(10^{-2})$ in the SM, the determination of this phase from (230) is affected by generic hadronic uncertainties of $\mathcal{O}(10\%)$, which may become an important issue for the LHC era. These uncertainties can be controlled with the help of flavour-symmetry arguments through the decay $B_d \to J/\psi\rho^0$ [156]. Needless to say, the big hope is that experiments will find a *sizeable* value of $\sin\phi_s$, which would immediately signal the presence of NP contributions to B_s^0–$\bar{B}_s^0$ mixing.

Other interesting aspects of the $B_s \to J/\psi\phi$ angular distribution are the determination of the width difference $\Delta\Gamma_s$ from untagged data samples [155] (for recent LHC feasibility studies, see Ref. [157]), and the extraction of $\cos\delta_f\cos\phi_s$ terms, where the δ_f are CP-conserving strong phases. If we fix the signs of $\cos\delta_f$ through factorization, we may extract the sign of $\cos\phi_s$, allowing an *unambiguous* determination of ϕ_s [95]. In this context, $B_s \to D_\pm\eta^{(\prime)}$, $D_\pm\phi$, ... decays offer also interesting methods [158], [159].

8.3 $B_s \to K^+K^-$

As can be seen from Fig. 14, the decay $B_d \to \pi^+\pi^-$ is related to the $B_s \to K^+K^-$ channel through an interchange of *all* down and strange quarks. Because of this feature, the U-spin flavour symmetry of strong interactions, which connects the down and strange quarks through $SU(2)$ transformations in the same manner as the ordinary isospin symmetry connects the down and up quarks, allows us to relate the hadronic $B_d \to \pi^+\pi^-$ parameters to their $B_s \to K^+K^-$ counterparts. It can then be shown that these quantities—and the angle γ of the UT—can be extracted from the measured CP asymmetries of the $B_d \to \pi^+\pi^-$, $B_s \to K^+K^-$ system [96]. Also other U-spin strategies were developed, using $B_{s(d)} \to J/\psi K_{\rm S}$ or $B_{d(s)} \to D^+_{d(s)}D^-_{d(s)}$ [87], $B_{d(s)} \to K^{0(*)}\bar{K}^{0(*)}$ [26], [156], $B_{(s)} \to \pi K$ [160], or $B_{s(d)} \to J/\psi\eta$ modes [161]. Since the $B_s \to K^+K^-$, $B_d \to \pi^+\pi^-$ system is particularly promising from an experimental point of view, thereby providing an interesting playground for CDF-II [18] and LHCb [19], [162], let us now have a closer look at the corresponding strategy [96].

8.3.1 Amplitude structure and CP asymmetries

If we follow Subsection 6.2, we may write the $B_d \to \pi^+\pi^-$, $B_s \to K^+K^-$ amplitudes as

$$A(B_d^0 \to \pi^+\pi^-) = \mathcal{C}\left[e^{i\gamma} - de^{i\theta}\right] \tag{231}$$

$$A(B_s^0 \to K^+K^-) = \left(\frac{\lambda}{1-\lambda^2/2}\right)\mathcal{C}'\left[e^{i\gamma} + \left(\frac{1-\lambda^2}{\lambda^2}\right)d'e^{i\theta'}\right], \tag{232}$$

where $de^{i\theta}$ was introduced in (154), $d'e^{i\theta'}$ is the $B_s \to K^+K^-$ counterpart of this quantity, and the overall normalization factors $\mathcal{C}$ and $\mathcal{C}'$ are CP-conserving strong amplitudes. Using these general parametrizations, we may write the corresponding CP-violating observables in the following generic form:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = {\rm fct}(d,\theta,\gamma), \qquad \mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = {\rm fct}(d,\theta,\gamma,\phi_d) \tag{233}$$

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_s \to K^+K^-) = {\rm fct}(d',\theta',\gamma), \quad \mathcal{A}_{\rm CP}^{\rm mix}(B_s \to K^+K^-) = {\rm fct}(d',\theta',\gamma,\phi_s). \tag{234}$$

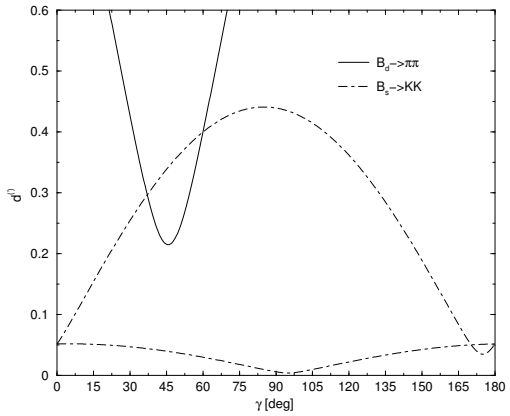


Fig. 23: The contours in the γ–$d^{(\prime)}$ plane for a specific example with $d = d' = 0.4$, $\theta = \theta' = 140°$, $\phi_d = 47°$, $\phi_s = 0°$, $\gamma = 60°$, corresponding to $\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-) = -0.30$, $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-) = +0.63$, $\mathcal{A}_{\rm CP}^{\rm dir}(B_s \to K^+K^-) = +0.16$ and $\mathcal{A}_{\rm CP}^{\rm mix}(B_s \to K^+K^-) = -0.17$.

The explicit expressions for the direct and mixing-induced CP asymmetries of $B_d \to \pi^+\pi^-$ are given in (158) and (159), respectively, whereas those for their $B_s \to K^+K^-$ counterparts can be found in Ref. [96]. Fortunately, these rather complicated expressions are not required for the following discussion.

8.3.2 Extraction of γ and hadronic parameters

As we saw in Subsection 6.1, ϕ_d can be extracted through the 'golden' mode $B_d \to J/\psi K_{\rm S}$, with the result in (151). On the other hand, ϕ_s can be assumed to be negligibly small in the SM, or can be fixed through $B_s \to J/\psi\phi$, as we discussed above. These experimental determinations work also in the presence of NP contributions to B_q^0–$\bar{B}_q^0$ mixing, as is obvious from the discussion in Subsection 6.4.

Looking at (233), we observe that a measurement of $\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-)$ and $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-)$ allows us to eliminate the strong phase θ, thereby yielding d as a function of γ in a *theoretically clean* way. In complete analogy, we may use the general parametrizations of the form in (234) to eliminate θ', and to determine d' in a *theoretically clean* manner as a function of γ from the measured values of $\mathcal{A}_{\rm CP}^{\rm dir}(B_s \to K^+K^-)$ and $\mathcal{A}_{\rm CP}^{\rm mix}(B_s \to K^+K^-)$. Since $B_d \to \pi^+\pi^-$ and $B_s \to K^+K^-$ are related to each other by interchanging all down and strange quarks, the U-spin flavour symmetry of strong interactions implies the following relations:

$$d' = d, \quad \theta' = \theta. \tag{235}$$

Applying the former, we may extract γ and d from the theoretically clean γ–d and γ–d' contours, which we have illustrated for a specific example in Fig. 23. As discussed in Ref. [96], it is also possible to resolve the twofold ambiguity for (γ, d) arising from the intersections of the solid and dot-dashed curves in Fig. 23. Moreover, we may determine θ and θ', thereby allowing an interesting internal consistency check of the second U-spin relation in (235).[11]

This strategy is very promising from an experimental point of view: at run II of the Tevatron and at the LHC, experimental accuracies for γ of $\mathcal{O}(10°)$ and $\mathcal{O}(1°)$, respectively, are expected [18], [162]. As far as the U-spin-breaking corrections to $d' = d$ are concerned, they enter the determination of γ through a relative shift of the γ–d and γ–d' contours; their impact on the extracted value of γ therefore depends on the form of these curves, which is fixed through the measured observables. In the examples discussed in Refs. [26], [96], as well as in the one shown in Fig. 23, the extracted value of γ would be

[11] Alternatively, we may eliminate d and d', and may then extract these parameters and γ through the relation $\theta' = \theta$.

very stable under such corrections. Let us also note that the U-spin relations in (235) appear to be quite robust, since the relevant form factors and decay constants cancel within factorization, so that they do not receive U-spin-breaking corrections in this approach [96]. On the other hand, the ratio $|\mathcal{C}'/\mathcal{C}|$, which equals 1 in the strict U-spin limit and enters the U-spin relation

$$\frac{\mathcal{A}_{\rm CP}^{\rm mix}(B_s \to K^+K^-)}{\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-)} = -\left|\frac{\mathcal{C}'}{\mathcal{C}}\right|^2 \left[\frac{{\rm BR}(B_d \to \pi^+\pi^-)}{{\rm BR}(B_s \to K^+K^-)}\right] \frac{\tau_{B_s}}{\tau_{B_d}}, \tag{236}$$

is affected by U-spin-breaking effects within factorization. An estimate of the corresponding form factors was recently performed in Ref. [163], and certain non-factorizable effects were addressed in Ref. [164].

8.3.3 Replacing $B_s \to K^+K^-$ by $B_d \to \pi^\mp K^\pm$

Since $B_s \to K^+K^-$ is not accessible at the e^+e^- B factories operating at the $\Upsilon(4S)$ resonance, we may not yet implement the strategy discussed above. However, as can easily be seen by looking at the corresponding Feynman diagrams, $B_s \to K^+K^-$ is related to $B_d \to \pi^\mp K^\pm$ through an interchange of spectator quarks. Consequently, we may approximately replace $B_s \to K^+K^-$ through $B_d \to \pi^\mp K^\pm$ in order to deal with the penguin problem in $B_d \to \pi^+\pi^-$ [165]. The utility of $B_d \to \pi^\mp K^\pm$ decays to control the penguin effects in $B_d \to \pi^+\pi^-$ was also emphasized in Ref. [98]. In order to explore the implications of the B-factory data, the following quantity plays a key role:

$$H = \frac{1}{\epsilon}\left(\frac{f_K}{f_\pi}\right)^2 \left[\frac{{\rm BR}(B_d \to \pi^+\pi^-)}{{\rm BR}(B_d \to \pi^\mp K^\pm)}\right] = 7.17 \pm 0.75. \tag{237}$$

Here $\epsilon \equiv \lambda^2/(1-\lambda^2)$, the ratio $f_K/f_\pi = 160/131$ describes factorizable $SU(3)$-breaking corrections, and the numerical value refers to the averages compiled in [86]. Applying (235), we obtain

$$H = \frac{1 - 2d\cos\theta\cos\gamma + d^2}{\epsilon^2 + 2\epsilon d\cos\theta\cos\gamma + d^2}. \tag{238}$$

If we now combine the CP asymmetries $\mathcal{A}_{\rm CP}^{\rm dir}(B_d \to \pi^+\pi^-)$ and $\mathcal{A}_{\rm CP}^{\rm mix}(B_d \to \pi^+\pi^-)$ with H, we have sufficient information available to determine γ, as well as d and θ [96], [165]. In practice, this can be done with the help of the expressions in (158), (159) and (238). A detailed discussion of this strategy was given in Refs. [47], [105], where also the impact of NP contributions to B^0_d–$\bar{B}^0_d$ mixing was explored. Using additional information from the $B \to \pi K$ analysis discussed in Subsection 7.2.3, the corresponding determination of γ was recently refined in [48], where in particular a twofold ambiguity for γ could be resolved, yielding

$$\gamma = (64.7^{+6.3}_{-6.9})^\circ, \tag{239}$$

which is in excellent agreement with the SM picture summarized in (168). If we complement this result with the experimental range for R_b and apply the simple relations

$$\bar{\rho} = R_b\cos\gamma, \quad \bar{\eta} = R_b\sin\gamma, \tag{240}$$

which follow straightforwardly from Fig. 4, we may also determine α and β:

$$\alpha = (93.6^{+10.3}_{-9.1})^\circ, \quad \beta = (21.7^{+2.5}_{-2.6})^\circ. \tag{241}$$

In Fig. 24, we compare these results with the allowed region for the apex of the UT that follows from the CKM fits, as implemented in [166].[12] Here the solid window corresponds to the range for γ in (239), whereas the dashed window indicates how the results change when the recently reported new Belle data [167] are used. Needless to note, the consistency of the overall picture is very remarkable.

[12] The small and large ellipses in Fig. 24 refer to the analyses of the SM and NP scenarios with MFV, respectively, as obtained in a recent update, see Ref. [82] of Ref. [166].

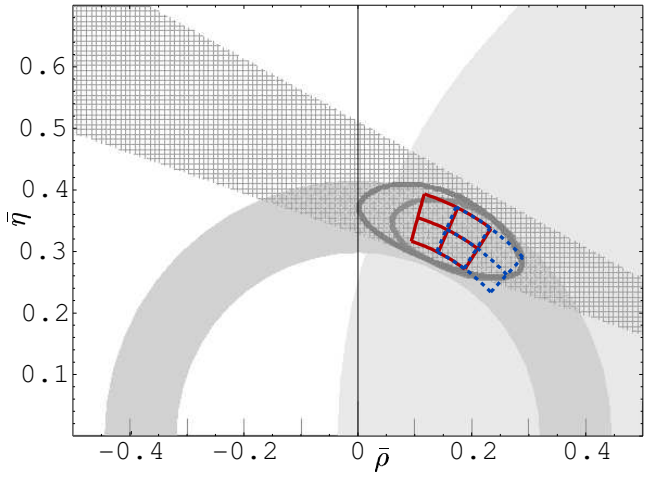


Fig. 24: Comparison of the determination of γ from the $B \to \pi\pi, \pi K$ data with the CKM fits, as discussed in the text.

In the analysis leading to (239) and (241), it has been assumed that $\phi_d \sim 47^\circ$, as in the SM. However, as discussed in Refs. [47], [105], it is interesting to consider also the second, unconventional solution of $\phi_d \sim 133^\circ$ in (151). There are simple relations to go from one solution to the other. In particular, if ϕ_d, γ, d and θ are solutions of (158), (159) and (238), then

$$\pi - \phi_d, \quad \pi - \gamma, \quad d, \quad \pi - \theta \tag{242}$$

are solutions as well. Consequently, (242) allows us to go easily from the $\phi_d \sim 47^\circ$ to the $\phi_d \sim 133^\circ$ case. Interestingly, for the value of θ in (169), we obtain $\cos\theta \sim -0.7 < 0$, having the same sign as in factorization, where $\cos\theta|_{\rm fact} = -1$. On the other hand, the value of θ corresponding to $\phi_d \sim 133^\circ$ yields $\cos\theta \sim +0.7 > 0$, i.e. the opposite sign, thereby disfavouring the $\phi_d \sim 133^\circ$ solution [48].

Let us finally note that the results for d and θ in (169) following from the $B \to \pi\pi$ analysis discussed in Subsection 6.2.2 allow us also to obtain SM predictions for the CP-violating $B_s \to K^+K^-$ observables with the help of (235) [48]:

$$\mathcal{A}_{\rm CP}^{\rm dir}(B_s \to K^+K^-)\Big|_{\rm SM} = 0.14^{+0.14}_{-0.09}, \quad \mathcal{A}_{\rm CP}^{\rm mix}(B_s \to K^+K^-)\Big|_{\rm SM} = -0.18^{+0.08}_{-0.07}. \tag{243}$$

On the other hand, the prediction of $\mathrm{BR}(B_s \to K^+K^-)$ requires information on the $SU(3)$-breaking form-factor ratios entering $|\mathcal{C}'/\mathcal{C}|$, where the estimates of Ref. [163] correspond to a branching ratio at the 3.5×10^{-5} level. It will be very interesting to see the first data for the $B_s \to K^+K^-$ channel from run II of the Tevatron, and to fully expoit its physics potential at LHCb and BTeV. The decay $B_s \to \pi^\pm K^\mp$ offers also various ways to complement the $B \to \pi\pi, \pi K$ strategy discussed in Subsection 7.2.3.

8.4 $B_s \to D_s^{(*)\pm} K^\mp$

Decays of the kind $B_s \to D_s^{(*)\pm} K^\mp, \ldots$ and their counterparts $B_d \to D^{(*)\pm}\pi^\mp, \ldots$ provide another important tool to explore CP violation [168], [169]. Since these transitions can be described on the same theoretical basis, we will consider them simultaneously in this subsection, following Ref. [170].

8.4.1 Basic features

It is convenient to write $B_s \to D_s^{(*)\pm} K^\mp, \ldots$ and $B_d \to D^{(*)\pm}\pi^\mp, \ldots$ decays generically as $B_q^0 \to D_q \bar{u}_q$, so that we may easily distinguish between the following cases:

– $q = s$: $D_s \in \{D_s^+, D_s^{*+}, \ldots\}$, $u_s \in \{K^+, K^{*+}, \ldots\}$.

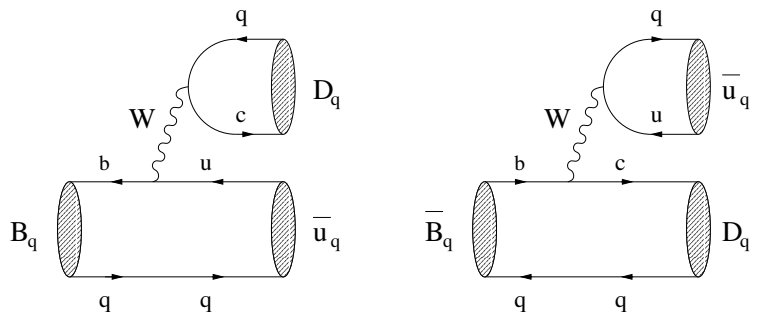


Fig. 25: Feynman diagrams contributing to $B_q^0 \to D_q \bar{u}_q$ and $\bar{B}_q^0 \to D_q \bar{u}_q$.

– $q = d$: $D_d \in \{D^+, D^{*+}, ...\}$, $u_d \in \{\pi^+, \rho^+, ...\}$.

In the discussion given below, we shall consider only those $B_q^0 \to D_q \bar{u}_q$ decays where at least one of the D_q, $\bar{u}_q$ states is a pseudoscalar meson. In the opposite case, for example $B_s^0 \to D_s^{*+} K^{*-}$, the extraction of weak phases would require a complicated angular analysis. If we look at Fig. 25, we observe that $B_q^0 \to D_q \bar{u}_q$ originates from colour-allowed tree topologies, and that also a $\bar{B}_q^0$ meson may decay into the same final-state $D_q \bar{u}_q$. The latter feature leads to interference effects between B_q^0–$\bar{B}_q^0$ mixing and decay processes, providing valuable information about the CP-violating phase $\phi_q + \gamma$.

8.4.2 Rate asymmetries

Let us first consider B_q decays into $D_q \bar{u}_q$. Since both a B_q^0 and a $\bar{B}_q^0$ meson may decay into this state, we obtain a time-dependent rate asymmetry of the following form:

$$\frac{\Gamma(B_q^0(t) \to D_q \bar{u}_q) - \Gamma(\bar{B}_q^0(t) \to D_q \bar{u}_q)}{\Gamma(B_q^0(t) \to D_q \bar{u}_q) + \Gamma(\bar{B}_q^0(t) \to D_q \bar{u}_q)} = \left[\frac{C(B_q \to D_q \bar{u}_q)\cos(\Delta M_q t) + S(B_q \to D_q \bar{u}_q)\sin(\Delta M_q t)}{\cosh(\Delta\Gamma_q t/2) - \mathcal{A}_{\Delta\Gamma}(B_q \to D_q \bar{u}_q)\sinh(\Delta\Gamma_q t/2)}\right], \tag{244}$$

having a structure that is completely analogous to the one of (126). Applying the formalism discussed in Section 5, we find that these observables are given by

$$C(B_q \to D_q \bar{u}_q) \equiv C_q = \frac{1 - |\xi_q|^2}{1 + |\xi_q|^2}, \quad S(B_q \to D_q \bar{u}_q) \equiv S_q = \frac{2\,\mathrm{Im}\,\xi_q}{1 + |\xi_q|^2}, \tag{245}$$

where

$$\xi_q \equiv -e^{-i\phi_q}\left[e^{i\phi_{\rm CP}(B_q)}\frac{A(\bar{B}_q^0 \to D_q \bar{u}_q)}{A(B_q^0 \to D_q \bar{u}_q)}\right] \tag{246}$$

measures the strength of the interference effects between the B_q^0–$\bar{B}_q^0$ mixing and decay processes.

If we take the Feynman diagrams shown in Fig. 25 into account and use an appropriate low-energy effective Hamiltonian of the kind discussed in Subsection 4.3.2, we may write

$$A(\bar{B}_q^0 \to D_q \bar{u}_q) = \langle \bar{u}_q D_q | \mathcal{H}_{\rm eff}(\bar{B}_q^0 \to D_q \bar{u}_q) | \bar{B}_q^0 \rangle = \frac{G_{\rm F}}{\sqrt{2}} \bar{v}_q \bar{M}_q, \tag{247}$$

where the hadronic matrix element

$$\bar{M}_q \equiv \langle \bar{u}_q D_q | \bar{\mathcal{O}}_1^q\, C_1(\mu) + \bar{\mathcal{O}}_2^q\, C_2(\mu) | \bar{B}_q^0 \rangle \tag{248}$$

involves the current–current operators

$$\bar{\mathcal{O}}_1^q \equiv (\bar{q}_\alpha u_\beta)_{\rm V\text{–}A}\,(\bar{c}_\beta b_\alpha)_{\rm V\text{–}A}\,, \quad \bar{\mathcal{O}}_2^q \equiv (\bar{q}_\alpha u_\alpha)_{\rm V\text{–}A}\,(\bar{c}_\beta b_\beta)_{\rm V\text{–}A}\,, \tag{249}$$

and the CKM factors $\bar{v}_q$ are given by

$$\bar{v}_s \equiv V_{us}^* V_{cb} = A\lambda^3, \quad \bar{v}_d \equiv V_{ud}^* V_{cb} = A\lambda^2(1-\lambda^2/2). \tag{250}$$

On the other hand, the $B_q^0 \to D_q \bar{u}_q$ decay amplitude takes the following form:

$$A(B_q^0 \to D_q \bar{u}_q) = \langle \bar{u}_q D_q | \mathcal{H}_{\rm eff}(B_q^0 \to D_q \bar{u}_q) | B_q^0 \rangle = \frac{G_{\rm F}}{\sqrt{2}} v_q^* \langle \bar{u}_q D_q | \mathcal{O}_1^{q\dagger}\, C_1(\mu) + \mathcal{O}_2^{q\dagger}\, C_2(\mu) | B_q^0 \rangle, \tag{251}$$

where we have to deal with the current–current operators

$$\mathcal{O}_1^q \equiv (\bar{q}_\alpha c_\beta)_{\rm V\text{–}A}\, (\bar{u}_\beta b_\alpha)_{\rm V\text{–}A}\,, \quad \mathcal{O}_2^q \equiv (\bar{q}_\alpha c_\alpha)_{\rm V\text{–}A}\, (\bar{u}_\beta b_\beta)_{\rm V\text{–}A}\,, \tag{252}$$

and the CKM factors v_q are defined as

$$v_s \equiv V_{cs}^* V_{ub} = A\lambda^3 R_b e^{-i\gamma}, \quad v_d \equiv V_{cd}^* V_{ub} = -\left(\frac{A\lambda^4 R_b}{1-\lambda^2/2}\right) e^{-i\gamma}. \tag{253}$$

If we introduce CP phases for the D_q and u_q mesons in analogy to (105), we obtain

$$(\mathcal{CP})|D_q \bar{u}_q\rangle = (-1)^L e^{i[\phi_{\rm CP}(D_q) - \phi_{\rm CP}(u_q)]} |\bar{D}_q u_q\rangle, \tag{254}$$

where L denotes the angular momentum of the $D_q \bar{u}_q$ state. Using now the relations $(\mathcal{CP})^\dagger(\mathcal{CP}) = \hat{1}$ and $(\mathcal{CP})\mathcal{O}_k^{q\dagger}(\mathcal{CP})^\dagger = \mathcal{O}_k^q$ as in Subsection 5.2, we may rewrite (251) as

$$A(B_q^0 \to D_q \bar{u}_q) = (-1)^L e^{i[\phi_{\rm CP}(B_q) - \phi_{\rm CP}(D_q) + \phi_{\rm CP}(u_q)]} \frac{G_{\rm F}}{\sqrt{2}} v_q^* M_q, \tag{255}$$

with

$$M_q \equiv \langle u_q \bar{D}_q | \mathcal{O}_1^q\, C_1(\mu) + \mathcal{O}_2^q\, C_2(\mu) | \bar{B}_q^0 \rangle. \tag{256}$$

An analogous calculation for the $\bar{B}_q^0 \to \bar{D}_q u_q$ and $B_q^0 \to \bar{D}_q u_q$ transitions yields

$$A(\bar{B}_q^0 \to \bar{D}_q u_q) = \frac{G_{\rm F}}{\sqrt{2}} v_q M_q \tag{257}$$

$$A(B_q^0 \to \bar{D}_q u_q) = (-1)^L e^{i[\phi_{\rm CP}(B_q) + \phi_{\rm CP}(D_q) - \phi_{\rm CP}(u_q)]} \frac{G_{\rm F}}{\sqrt{2}} \bar{v}_q^* \bar{M}_q, \tag{258}$$

where the same hadronic matrix elements as in the $B_q^0 \to D_q \bar{u}_q$ and $\bar{B}_q^0 \to D_q \bar{u}_q$ modes arise.

If we now insert (247) and (255) into (246), we observe that the convention-dependent phase $\phi_{\rm CP}(B_q)$ is cancelled through the amplitude ratio, and arrive at

$$\xi_q = -(-1)^L e^{-i(\phi_q + \gamma)} \left[\frac{1}{x_q e^{i\delta_q}}\right], \tag{259}$$

where

$$x_s \equiv R_b a_s, \quad x_d \equiv -\left(\frac{\lambda^2 R_b}{1-\lambda^2}\right) a_d, \tag{260}$$

with

$$a_q e^{i\delta_q} \equiv e^{-i[\phi_{\rm CP}(D_q) - \phi_{\rm CP}(u_q)]} \frac{M_q}{\bar{M}_q}. \tag{261}$$

It should be noted that the convention-dependent phases $\phi_{\rm CP}(D_q)$ and $\phi_{\rm CP}(u_q)$ in (261) are cancelled through the ratio of hadronic matrix elements, so that $a_q e^{i\delta_q}$ is actually a physical observable (this is shown explicitly in Ref. [170]). Applying now (245), we finally arrive at

$$C_q = -\left[\frac{1-x_q^2}{1+x_q^2}\right], \quad S_q = (-1)^L \left[\frac{2\, x_q \sin(\phi_q + \gamma + \delta_q)}{1+x_q^2}\right]. \tag{262}$$

An analogous calculation for the decays into the CP-conjugate final-state $\bar{D}_q u_q$ yields

$$\bar{\xi}_q = -e^{-i\phi_q}\left[e^{i\phi_{\rm CP}(B_q)}\frac{A(\bar{B}_q^0\to\bar{D}_q u_q)}{A(B_q^0\to\bar{D}_q u_q)}\right] = -(-1)^L e^{-i(\phi_q+\gamma)}\left[x_q e^{i\delta_q}\right], \tag{263}$$

which implies

$$\bar{C}_q = +\left[\frac{1-x_q^2}{1+x_q^2}\right], \quad \bar{S}_q = (-1)^L\left[\frac{2\,x_q\sin(\phi_q+\gamma-\delta_q)}{1+x_q^2}\right], \tag{264}$$

where $\bar{C}_q \equiv C(B_q\to\bar{D}_q u_q)$ and $\bar{S}_q \equiv S(B_q\to\bar{D}_q u_q)$. Note that $\bar{\xi}_q$ and ξ_q satisfy the relation

$$\bar{\xi}_q \times \xi_q = e^{-i2(\phi_q+\gamma)}, \tag{265}$$

where the hadronic parameter $x_q e^{i\delta_q}$ *cancels*. Consequently, we may extract $\phi_q+\gamma$ in a *theoretically clean* way from the corresponding observables.

8.4.3 Conventional extraction of $\phi_q+\gamma$

It is convenient to introduce the following combinations of observables:

$$\langle C_q\rangle_+ \equiv \frac{\bar{C}_q + C_q}{2} = 0 \tag{266}$$

$$\langle C_q\rangle_- \equiv \frac{\bar{C}_q - C_q}{2} = \frac{1-x_q^2}{1+x_q^2} \tag{267}$$

$$\langle S_q\rangle_+ \equiv \frac{\bar{S}_q + S_q}{2} = +(-1)^L\left[\frac{2\,x_q\cos\delta_q}{1+x_q^2}\right]\sin(\phi_q+\gamma) \tag{268}$$

$$\langle S_q\rangle_- \equiv \frac{\bar{S}_q - S_q}{2} = -(-1)^L\left[\frac{2\,x_q\sin\delta_q}{1+x_q^2}\right]\cos(\phi_q+\gamma). \tag{269}$$

We observe that (267) allows us—in principle—to determine x_q from $\langle C_q\rangle_-$. However, to this end, terms entering at the x_q^2 level have to be resolved experimentally. In the case of $q=s$, we have $x_s = \mathcal{O}(R_b)$, implying $x_s^2 = \mathcal{O}(0.16)$, so that this may actually be possible, although challenging [168]. On the other hand, $x_d = \mathcal{O}(-\lambda^2 R_b)$ is doubly Cabibbo-suppressed. Although it should be possible to resolve terms of $\mathcal{O}(x_d)$, this will be impossible for the vanishingly small $x_d^2 = \mathcal{O}(0.0004)$ terms, so that alternative approaches to fix x_d are required [169].

In contrast to the observables associated with the $\cos(\Delta M_q t)$ terms, the mixing-induced observables entering the rate asymmetries with $\sin(\Delta M_q t)$ provide information on $\phi_q+\gamma$. Let us now assume that x_q is known. We may then consider

$$s_+ \equiv (-1)^L\left[\frac{1+x_q^2}{2x_q}\right]\langle S_q\rangle_+ = +\cos\delta_q\sin(\phi_q+\gamma) \tag{270}$$

$$s_- \equiv (-1)^L\left[\frac{1+x_q^2}{2x_q}\right]\langle S_q\rangle_- = -\sin\delta_q\cos(\phi_q+\gamma), \tag{271}$$

yielding

$$\sin^2(\phi_q+\gamma) = \frac{1}{2}\left[(1+s_+^2-s_-^2) \pm \sqrt{(1+s_+^2-s_-^2)^2 - 4s_+^2}\right]. \tag{272}$$

This expression implies an eightfold solution for $\phi_q+\gamma$. If we fix the sign of $\cos\delta_q$ with the help of factorization, a fourfold discrete ambiguity emerges. Since we may determine ϕ_d and ϕ_s through analyses of $B_d\to J/\psi K_{\rm S}$ and $B_s\to J/\psi\phi$ decays, respectively, we may extract γ from $\phi_q+\gamma$.

8.4.4 New strategies and recent developments

Let us now discuss new strategies to explore the $B_q \to D_q \bar{u}_q$ modes [170]. If the width difference $\Delta\Gamma_s$ is sizeable, the time-dependent untagged rates [see (229)]

$$\langle\Gamma(B_q(t) \to D_q\bar{u}_q)\rangle = \langle\Gamma(B_q \to D_q\bar{u}_q)\rangle\,[\cosh(\Delta\Gamma_q t/2) - \mathcal{A}_{\Delta\Gamma}(B_q \to D_q\overline{u}_q)\,\sinh(\Delta\Gamma_q t/2)]\,e^{-\Gamma_q t} \tag{273}$$

and their CP conjugates provide $\mathcal{A}_{\Delta\Gamma}(B_s \to D_s\bar{u}_s) \equiv \mathcal{A}_{\Delta\Gamma_{\rm s}}$ and $\mathcal{A}_{\Delta\Gamma}(B_s \to \bar{D}_s u_s) \equiv \bar{\mathcal{A}}_{\Delta\Gamma_{\rm s}}$. It can be shown that these 'untagged' observables can be combined with their 'tagged' counterparts $\langle S_s\rangle_\pm$ in the form of the following simple relation:

$$\tan(\phi_s + \gamma) = -\left[\frac{\langle S_s\rangle_+}{\langle\mathcal{A}_{\Delta\Gamma_{\rm s}}\rangle_+}\right] = +\left[\frac{\langle\mathcal{A}_{\Delta\Gamma_{\rm s}}\rangle_-}{\langle S_s\rangle_-}\right], \tag{274}$$

where $\langle\mathcal{A}_{\Delta\Gamma_{\rm s}}\rangle_+$ and $\langle\mathcal{A}_{\Delta\Gamma_{\rm s}}\rangle_-$ are defined in analogy to (268) and (269), respectively. Obviously, (274) offers an elegant extraction of $\phi_s+\gamma$, up to a twofold ambiguity. If we fix again the sign of $\cos\delta_q$ through factorization, we may determine $\phi_s + \gamma$ in an *unambiguous* manner, which should be compared with the fourfold ambiguity arising in this case from (272). In particular, we may decide whether $\gamma \in [0^\circ, 180^\circ]$, as in the SM, or $\gamma \in [180^\circ, 360^\circ]$. Another important advantage of (274) is that we do *not* have to rely on the resolution of $\mathcal{O}(x_s^2)$ terms, as $\langle S_s\rangle_\pm$ and $\langle\mathcal{A}_{\Delta\Gamma_{\rm s}}\rangle_\pm$ are both proportional to x_s. On the other hand, we need a sizeable value of $\Delta\Gamma_s$. Measurements of untagged rates are also very useful in the case of a vanishingly small $\Delta\Gamma_q$, since the 'unevolved' (i.e. time-independent) untagged rates in (273) offer various interesting strategies to determine x_q from the ratio of $\langle\Gamma(B_q \to D_q\bar{u}_q)\rangle + \langle\Gamma(B_q \to \bar{D}_q u_q)\rangle$ to CP-averaged rates of appropriate $B^\pm$ or flavour-specific B_q decays.

If we keep the hadronic parameter x_q and the associated strong phase δ_q as 'unknown', free parameters in the expressions for the $\langle S_q\rangle_\pm$, we may derive the relations

$$|\sin(\phi_q + \gamma)| \geq |\langle S_q\rangle_+|, \quad |\cos(\phi_q + \gamma)| \geq |\langle S_q\rangle_-|, \tag{275}$$

which can straightforwardly be converted into bounds on $\phi_q + \gamma$. If x_q is known, stronger constraints are implied by

$$|\sin(\phi_q + \gamma)| \geq |s_+|, \quad |\cos(\phi_q + \gamma)| \geq |s_-|. \tag{276}$$

Once s_+ and s_- are known, we may of course determine $\phi_q + \gamma$ through the 'conventional' approach, using (272). However, the bounds following from (276) provide essentially the same information and are much simpler to implement. Moreover, as discussed in detail in Ref. [170] for several examples, the bounds following from the B_s and B_d modes may be highly complementary, thereby providing particularly narrow, theoretically clean ranges for γ. Whereas the B_s decays are not yet accessible, first results for the $B_d \to D^{(*)\pm}\pi^\mp$ modes obtained by BaBar give $|\sin(\phi_d + \gamma)| > 0.87\ (0.58)$ at the 68% (95%) C.L. [171]. Looking at (268), we observe that we may extract the sign of $\sin(\phi_q + \gamma)$ from $\langle S_q\rangle_+$ if we assume that the sign of $\cos\delta_q$ is as in factorization. To this end, the factor $(-1)^L$ has to be properly taken into account. The information on the sign of $\sin(\phi_d + \gamma)$ is very useful, as it allows us to distinguish directly between the two solutions for (ϕ_d, γ) discussed in Subsection 8.3.3. If we apply (242), the analysis of CP violation in $B_d \to \pi^+\pi^-$ gives $(\phi_d, \gamma) \sim (47^\circ, 65^\circ)$ or $(133^\circ, 115^\circ)$ [47], [105], corresponding to $\sin(\phi_q + \gamma) \sim +0.9$ or -0.9, respectively. The BaBar analysis favours the former case [170], i.e. the picture of the SM, in accordance with the discussion after (242). The exploration of $B_d \to D^{(*)\pm}\pi^\mp$ modes is also in progress at Belle [172]. Unfortunately, the current Belle results for (fully reconstructed) $B_d \to D^{(*)\pm}\pi^\mp$ decays favour the sign opposite to the one obtained by BaBar (see also Ref. [86]), so that the experimental picture is not yet conclusive.

Let us now further exploit the complementarity between the $B_s^0 \to D_s^{(*)+}K^-$ and $B_d^0 \to D^{(*)+}\pi^-$ modes. If we look at their decay topologies, we observe that these channels are related to each other through an interchange of all down and strange quarks. Consequently, the U-spin flavour symmetry of

strong interactions implies $a_s = a_d$ and $\delta_s = \delta_d$. There are various possibilities to implement these relations. A particularly simple picture emerges if we assume that $a_s = a_d$ *and* $\delta_s = \delta_d$, which yields

$$\tan\gamma = -\left[\frac{\sin\phi_d - S\sin\phi_s}{\cos\phi_d - S\cos\phi_s}\right] \stackrel{\phi_s=0^\circ}{=} -\left[\frac{\sin\phi_d}{\cos\phi_d - S}\right]. \tag{277}$$

Here we have introduced

$$S \equiv -R\left[\frac{\langle S_d\rangle_+}{\langle S_s\rangle_+}\right] \tag{278}$$

with

$$R \equiv \left(\frac{1-\lambda^2}{\lambda^2}\right)\left[\frac{1}{1+x_s^2}\right], \tag{279}$$

which can be fixed from untagged B_s rates through

$$R = \left(\frac{f_K}{f_\pi}\right)^2\left[\frac{\Gamma(\bar{B}_s^0 \to D_s^{(*)+}\pi^-) + \Gamma(B_s^0 \to D_s^{(*)-}\pi^+)}{\langle\Gamma(B_s \to D_s^{(*)+}K^-)\rangle + \langle\Gamma(B_s \to D_s^{(*)-}K^+)\rangle}\right]. \tag{280}$$

Alternatively, we may *only* assume that $\delta_s = \delta_d$ *or* that $a_s = a_d$. Apart from features related to multiple discrete ambiguities, the most important advantage with respect to the 'conventional' approach is that the experimental resolution of the x_q^2 terms is not required. In particular, x_d does *not* have to be fixed, and x_s may only enter through a $1 + x_s^2$ correction, which can straightforwardly be determined through untagged B_s rate measurements. In the most refined implementation of this strategy, the measurement of x_d/x_s would *only* be interesting for the inclusion of U-spin-breaking effects in a_d/a_s.

9 RARE DECAYS

9.1 General features and impact of New Physics in models with minimal flavour violation

In order to complement the exploration of flavour physics through the CP-violating phenomena discussed above, also various rare decays of B and K mesons offer very interesting strategies. As we have already noted, by 'rare' decays we mean transitions that do *not* arise at the tree level in the SM, but may originate through loop effects. Consequently, rare B decays are mediated by FCNC processes of the kind $\bar{b} \to \bar{s}$ or $\bar{b} \to \bar{d}$, whereas rare K decays originate from their $\bar{s} \to \bar{d}$ counterparts. Prominent examples of rare B decays are the following exclusive decay modes:

- $B \to K^*\gamma$, $B \to \rho\gamma$, ...
- $B \to K\mu^+\mu^-$, $B \to \pi\mu^+\mu^-$, ...
- $B_{s,d} \to \mu^+\mu^-$.

While the $B_{s,d} \to \mu^+\mu^-$ transitions are very clean, the former two decay classes suffer from theoretical uncertainties that are related to hadronic form factors and long-distance contributions. On the other hand, the hadronic uncertainties are much smaller in the corresponding inclusive decays, $B \to X_{s,d}\gamma$ and $B \to X_{s,d}\mu^+\mu^-$, which are therefore more promising from the theoretical point of view, but are unfortunately more difficult to measure; the cleanest rare B decays are given by $B \to X_{s,d}\nu\bar{\nu}$ processes. Let us note that a tremendous amount of work went into the calculation of the branching ratio of the prominent $B \to X_s\gamma$ channel (for an overview, see [173]); the agreement of the experimental value with the SM expectation implies important constraints for the allowed parameter space of popular NP scenarios. The phenomenology of the kaon system includes also interesting rare decays:

- $K_{\rm L} \to \mu^+\mu^-$
- $K_{\rm L} \to \pi^0 e^+ e^-$
- $K_{\rm L} \to \pi^0\nu\bar{\nu}$, $K^+ \to \pi^+\nu\bar{\nu}$,

where the 'golden' modes are given by the $K \to \pi\nu\bar{\nu}$ processes, which are essentially theoretically clean, as we have already noted in Subsection 3.2.

In order to deal with rare decays theoretically, appropriate low-energy effective Hamiltonians are used, in analogy to the analysis of non-leptonic B decays. The structure of the corresponding transition amplitudes is similar to the one of (77), i.e. the short-distance physics is described by perturbatively calculable Wilson coefficient functions, whereas the long-distance dynamics is encoded in non-perturbative hadronic matrix elements of local operators. It is useful to follow Refs. [174], [175], and to rewrite the rare-decay implementation of (77) as

$$A(\text{decay}) = P_0(\text{decay}) + \sum_r P_r(\text{decay}) F_r(v). \tag{281}$$

For the derivation of this expression, we choose $\mu = \mu_0 = \mathcal{O}(M_W)$, and rewrite the corresponding Wilson coefficients $C_k(\mu_0)$ as linear combinations of 'master functions' $F_r(v)$, which follow from the evaluation of penguin and box diagrams with heavy particle exchanges. Expression (281) applies not only to the SM, but also to NP scenarios with MFV (see Subsection 6.4.4), where the parameters involved are collectively denoted by v. In the SM, the functions $F_r(v)$ reduce to the well-known Inami–Lim functions [81], with $v = x_t = m_t^2/M_W^2$. The term P_0 summarizes the contributions that originate from light internal quarks, such as the charm and up quarks, and the sum takes the remaining contributions into account. For a detailed discussion of this formalism and the general features of the P_0, P_r and F_r, we refer the reader to Ref. [122]. Let us here just emphasize the following important points:

- The $F_r(v)$ are *process-independent, universal* functions that depend on the particular model considered. NP enters the decay amplitudes only through these functions.
- The P_0 and P_r are *process-dependent* quantities. In particular, they depend on the hadronic matrix elements of the operators Q_k.

In models with MFV, the set of the $F_r(v)$ consists of seven functions

$$S(v),\ X(v),\ Y(v),\ Z(v),\ E(v),\ D'(v),\ E'(v), \tag{282}$$

which are discussed in detail in Ref. [122]. In (184), we already encountered one of them, the function $S(v)$, which governs B_q^0–$\bar{B}_q^0$ and K^0–$\bar{K}^0$ mixing; below, we shall come across $X(v)$ and $Y(v)$, which characterize rare K, B decays with $\nu\bar{\nu}$ and $\ell^+\ell^-$ in the final states, respectively. The important property of the functions in (282) is that they do not—within the framework of MFV—contain complex phases, so that the CP-violating effects are governed *entirely* by the KM phase hiding in the parameters P_r.

For detailed discussions of the many interesting aspects of rare B and K decays and recent developments, we refer the reader to Refs. [21], [22], [173], [176]. Let us here choose $B_{s,d} \to \mu^+\mu^-$ and $K \to \pi\nu\bar{\nu}$ processes as representative examples, which are particularly clean from the theoretical point of view; the former channels are also an important element of the B-physics programme of the LHC [19]. Finally, we shall illustrate the impact of NP that does *not* belong to the class of MFV models on rare decays. To this end, we consider a NP scenario that is suggested by the '$B \to \pi K$ puzzle' discussed in Subsection 7.2.3.

9.2 $B_{s,d} \to \mu^+\mu^-$

As can be seen in Fig. 26, within the framework of the SM, the decays $B_{s,d} \to \mu^+\mu^-$ originate from Z^0 penguins and box diagrams. These transitions belong to the cleanest modes in the category of rare B decays, since they involve only the hadronic matrix element of a quark current between a B_q-meson and the vacuum state, i.e. the decay constant f_{B_q} that we introduced in (56), NLO QCD corrections were calculated, and long-distance contributions are expected to play a negligible role [177]. The low-energy effective Hamiltonian describing $B_q \to \mu^+\mu^-$ decays is given as follows ($q \in \{s, d\}$):

$$\mathcal{H}_{\text{eff}} = -\frac{G_{\text{F}}}{\sqrt{2}} \left[\frac{\alpha}{2\pi \sin^2\Theta_{\text{W}}}\right] V_{tb}^* V_{tq} \eta_Y Y_0(x_t) (\bar{b}q)_{\text{V-A}} (\bar{\mu}\mu)_{\text{V-A}} + \text{h.c.}, \tag{283}$$

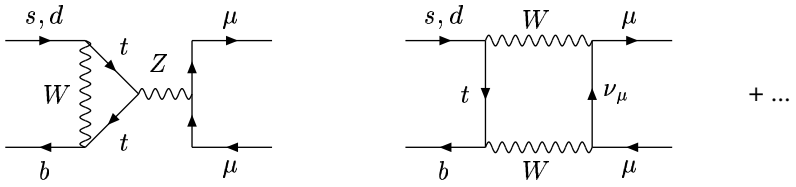


Fig. 26: Decay processes contributing to $B_{s,d} \to \mu^+\mu^-$ in the SM.

where α denotes the QED coupling and $\Theta_{\rm W}$ is the Weinberg angle. Here the short-distance physics is described by

$$Y(x_t) = \eta_Y Y_0(x_t), \tag{284}$$

where $\eta_Y = 1.012$ is a perturbative QCD correction factor [177], [178], [179], and $Y_0(x_t)$, which is another Inami–Lim function [81], describes the top-quark mass dependence of the Feynman diagrams shown in Fig. 26. In the SM, we may write $Y_0(x_t)$—to a very good approximation—as follows [122]:

$$Y_0(x_t) = 0.98 \times \left[\frac{m_t}{167\,{\rm GeV}}\right]^{1.56}. \tag{285}$$

We observe that the matrix element of (283) between a $\langle\mu^-\mu^+|$ final state and a $|B_q\rangle$ initial state indeed involves the decay constant f_{B_q}. The corresponding SM branching ratios then take the following form [37]:

$${\rm BR}(B_s \to \mu^+\mu^-) = 4.1 \times 10^{-9} \left[\frac{f_{B_s}}{0.24\,{\rm GeV}}\right]^2 \left[\frac{|V_{ts}|}{0.040}\right]^2 \left[\frac{\tau_{B_s}}{1.5\,{\rm ps}}\right] \left[\frac{m_t}{167\,{\rm GeV}}\right]^{3.12} \tag{286}$$

$${\rm BR}(B_d \to \mu^+\mu^-) = 1.1 \times 10^{-10} \left[\frac{f_{B_d}}{0.20\,{\rm GeV}}\right]^2 \left[\frac{|V_{td}|}{0.008}\right]^2 \left[\frac{\tau_{B_d}}{1.5\,{\rm ps}}\right] \left[\frac{m_t}{167\,{\rm GeV}}\right]^{3.12}, \tag{287}$$

which should be compared with the experimental 90% C.L. bounds

$${\rm BR}(B_s \to \mu^+\mu^-) < 5.8 \times 10^{-7}, \quad {\rm BR}(B_d \to \mu^+\mu^-) < 1.5\,(1.6) \times 10^{-7} \tag{288}$$

obtained by the CDF (Belle) collaboration [180]. Looking at (286) and (287), we see that a measurement of these branching ratios would allow clean determinations of $|V_{ts}|$ and $|V_{td}|$, respectively, provided the non-perturbative decay constants f_{B_s} and f_{B_d} were known reliably. The current status following from lattice QCD studies is given as follows [41]:

$$f_{B_d} = (203 \pm 27^{+0}_{-20})\,{\rm MeV}, \quad f_{B_s} = (238 \pm 31)\,{\rm MeV}; \tag{289}$$

similar results were obtained with the help of QCD sum rules [150]. If we consider the ratio

$$\frac{{\rm BR}(B_d \to \mu^+\mu^-)}{{\rm BR}(B_s \to \mu^+\mu^-)} = \left[\frac{\tau_{B_d}}{\tau_{B_s}}\right] \left[\frac{M_{B_d}}{M_{B_s}}\right] \left[\frac{f_{B_d}}{f_{B_s}}\right]^2 \left|\frac{V_{td}}{V_{ts}}\right|^2, \tag{290}$$

these parameters enter only in the form of the following $SU(3)$-breaking ratio (see also (227)):

$$\frac{f_{B_s}}{f_{B_d}} = 1.18 \pm 0.04^{+0.12}_{-0}. \tag{291}$$

Using now (224), the relation in (290) allows a determination of the side R_t of the UT. On the other hand, we may also write (see (222))

$$\frac{\Delta M_d}{\Delta M_s} = \left[\frac{M_{B_d}}{M_{B_s}}\right] \left[\frac{\hat{B}_{B_d}}{\hat{B}_{B_s}}\right] \left[\frac{f_{B_d}}{f_{B_s}}\right]^2 \left|\frac{V_{td}}{V_{ts}}\right|^2, \tag{292}$$

allowing us to fix R_t with the help of (225). Consequently, (290) and (292) provide complementary determinations of the UT side R_t. Moreover, these expressions imply also the following relation:

$$\frac{\mathrm{BR}(B_s \to \mu^+\mu^-)}{\mathrm{BR}(B_d \to \mu^+\mu^-)} = \left[\frac{\tau_{B_s}}{\tau_{B_d}}\right]\left[\frac{\hat{B}_{B_d}}{\hat{B}_{B_s}}\right]\left[\frac{\Delta M_s}{\Delta M_d}\right], \tag{293}$$

which suffers from theoretical uncertainties that are smaller than those affecting (290) and (292), since the dependence on $(f_{B_d}/f_{B_s})^2$ cancels, and $\hat{B}_{B_d}/\hat{B}_{B_s} = 1$ up to tiny $SU(3)$-breaking corrections [181]. In particular, QCD lattice simulations give the following numbers [41]:

$$\hat{B}_{B_d} = 1.34 \pm 0.12, \quad \hat{B}_{B_s} = 1.34 \pm 0.12, \quad \frac{\hat{B}_{B_s}}{\hat{B}_{B_d}} = 1.00 \pm 0.03. \tag{294}$$

Moreover, we may also use the (future) experimental data for $\Delta M_{(s)d}$ to reduce the hadronic uncertainties in the SM predictions for the $B_q \to \mu^+\mu^-$ branching ratios [181], yielding

$$\mathrm{BR}(B_s \to \mu^+\mu^-)\big|_{\mathrm{SM}} = (3.42 \pm 0.53) \times \left[\frac{\Delta M_s}{18.0\,\mathrm{ps}^{-1}}\right] \times 10^{-9} \tag{295}$$

$$\mathrm{BR}(B_d \to \mu^+\mu^-)\big|_{\mathrm{SM}} = (1.00 \pm 0.14) \times 10^{-10}. \tag{296}$$

Since these branching ratios are very small, we could only hope to observe the $B_q \to \mu^+\mu^-$ decays at the LHC, should they actually be governed by their SM contributions [19]. However, as these transitions are mediated by rare FCNC processes, they are sensitive probes for NP. In particular, as was recently reviewed in Ref. [182], the $B_q \to \mu^+\mu^-$ branching ratios may be dramatically enhanced in specific NP (SUSY) scenarios. Should this actually be the case, these decays may be seen at run II of the Tevatron, and the e^+e^- B factories could observe $B_d \to \mu^+\mu^-$. In the case of models with MFV, we just have to make the replacement

$$Y(x_t) \to Y(v) \tag{297}$$

in order to take the NP contributions to the $B_q \to \mu^+\mu^-$ decays into account. In particular, the *same* $Y(v)$ enters the $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ channels (see (282)). In analogy, the *same* generalized function $S(v)$ governs the mass differences ΔM_s and ΔM_d, as we have seen in Subsection 6.4.4. Consequently, within MFV scenarios, the NP effects cancel in (290), (292) and (293), where in particular the latter relation offers an interesting test of this picture.

9.3 $K \to \pi\nu\bar{\nu}$

As we discussed in Subsection 3.2, $K \to \pi\nu\bar{\nu}$ decays originate from Z^0 penguins and box diagrams. Let us first have a closer look at the charged mode $K^+ \to \pi^+\nu\bar{\nu}$. The low-energy effective Hamiltonian describing this decay is given as follows [37]:

$$\mathcal{H}_{\mathrm{eff}} = \frac{G_{\mathrm{F}}}{\sqrt{2}}\left[\frac{\alpha}{2\pi\sin^2\Theta_{\mathrm{W}}}\right]\sum_{\ell=e,\mu,\tau}\left[\lambda_c X_{\mathrm{NL}}^{\ell} + \lambda_t X(x_t)\right](\bar{s}d)_{\mathrm{V-A}}(\bar{\nu}_\ell\nu_\ell)_{\mathrm{V-A}}, \tag{298}$$

where

$$\lambda_c \equiv V_{cs}^* V_{cd} = -\lambda\left(1 - \frac{\lambda^2}{2}\right) \tag{299}$$

and $\lambda_t \equiv V_{ts}^* V_{td}$ with

$$\mathrm{Im}\,\lambda_t = \eta A^2\lambda^5, \quad \mathrm{Re}\,\lambda_t = -\left(1 - \frac{\lambda^2}{2}\right)A^2\lambda^5(1-\bar{\rho}) \tag{300}$$

are CKM factors, and

$$X(x_t) = \eta_X X_0(x_t) \tag{301}$$

describes the top-quark mass dependence originating from the Z^0 penguin and box diagrams, where $X_0(x_t)$ is another Inam–Lim function [81], and $\eta_X = 0.994$ is a perturbative NLO QCD correction factor [177], [178], [179], [183]. Within the SM, we may write $X_0(x_t)$—to a very good approximation—as follows [122]:

$$X_0(x_t) = 1.53 \times \left[\frac{m_t}{167\,\mathrm{GeV}}\right]^{1.15}. \tag{302}$$

The counterpart of $X(x_t)$ in the charm sector is given by $X_{\rm NL}^{\ell}$. For the analysis of $K^+ \to \pi^+\nu\bar{\nu}$, the following combination is relevant:[13]

$$P_c(\nu\bar{\nu}) = \frac{1}{\lambda^4}\left[\frac{2}{3}X_{\rm NL}^{e} + \frac{1}{3}X_{\rm NL}^{\tau}\right] = 0.39 \pm 0.06. \tag{303}$$

If we calculate the matrix element of (298) between the $\langle\bar{\nu}\nu\pi^+|$ final state and the $|K^+\rangle$ initial state, we encounter a hadronic matrix element of the $(\bar{s}d)_{\rm V-A}$ current that can be extracted—with the help of the isospin flavour symmetry of strong interactions—from the semileptonic decay $K^+ \to \pi^0 e^+\nu$, which is a tree decay that is described by the following Hamiltonian [37]:

$$\mathcal{H}_{\rm eff}(K^+ \to \pi^0 e^+\nu) = \frac{G_{\rm F}}{\sqrt{2}}V_{us}^*(\bar{s}u)_{\rm V-A}(\bar{\nu}_e e)_{\rm V-A}. \tag{304}$$

Using the isospin relation

$$\langle\pi^+|(\bar{s}d)_{\rm V-A}|K^+\rangle = \sqrt{2}\langle\pi^0|(\bar{s}u)_{\rm V-A}|K^+\rangle, \tag{305}$$

and neglecting the phase-space differences due to $M_{\pi^+} \neq M_{\pi^0}$ and $M_e \neq 0$, we obtain

$$\frac{\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})}{\mathrm{BR}(K^+ \to \pi^0 e^+\nu)} = \frac{\alpha^2}{|V_{us}|^2 2\pi^2\sin^4\Theta_{\rm W}}\sum_{\ell=e,\mu,\tau}|\lambda_c X_{\rm NL}^{\ell} + \lambda_t X(x_t)|^2. \tag{306}$$

Consequently, we may determine the hadronic matrix element relevant to the rare decay $K^+ \to \pi^+\nu\bar{\nu}$ through the experimental data for the (non-rare) decay $K^+ \to \pi^0 e^+\nu$. Because of this important feature, $K^+ \to \pi^+\nu\bar{\nu}$ is a very *clean* decay.

It is useful to write the $K^+ \to \pi^+\nu\bar{\nu}$ branching ratio as

$$B_1 \equiv \frac{1}{\kappa_+}\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu}), \tag{307}$$

with

$$\kappa_+ = r_{K^+}\left[\frac{3\alpha^2\mathrm{BR}(K^+ \to \pi^0 e^+\nu)}{2\pi^2\sin^4\Theta_{\rm W}}\right]\lambda^8 = 4.78 \times 10^{-11}, \tag{308}$$

where $r_{K^+} = 0.901$ describes the isospin-breaking corrections that arise in relating $K^+ \to \pi^+\nu\bar{\nu}$ to $K^+ \to \pi^0 e^+\nu$. Let us now consider the general MFV case, where

$$X(x_t) \to X(v). \tag{309}$$

The 'reduced' $K^+ \to \pi^+\nu\bar{\nu}$ branching ratio B_1 can then be expressed as follows [126]:

$$B_1 = \left[\frac{\mathrm{Im}\lambda_t}{\lambda^5}|X(v)|\right]^2 + \left[\frac{\mathrm{Re}\lambda_c}{\lambda}\mathrm{sgn}(X(v))P_c(\nu\bar{\nu}) + \frac{\mathrm{Re}\lambda_t}{\lambda^5}|X(v)|\right]^2; \tag{310}$$

[13] The small numerical difference of $P_c(\nu\bar{\nu})$ with respect to the value given in Ref. [37], where $\lambda = 0.2205$ was used, is related to the very recent value of $\lambda = 0.2240$ [41]. A similar comment applies to the quantities κ_+ and $\kappa_{\rm L}$, to be introduced below.

the corresponding SM prediction following from the very recent update in Ref. [48] was given in (49). It is now an easy exercise to show that the measured $K^+ \to \pi^+ \nu\bar{\nu}$ branching ratio determines an ellipse in the $\bar{\rho}$–$\bar{\eta}$ plane,

$$\left(\frac{\bar{\rho}-\rho_0}{\bar{\rho}_1}\right)^2 + \left(\frac{\bar{\eta}}{\bar{\eta}_1}\right)^2 = 1, \tag{311}$$

centred at $(\rho_0, 0)$ with

$$\rho_0 = 1 + \mathrm{sgn}(X(v))\frac{P_c(\nu\bar{\nu})}{A^2|X(v)|}, \tag{312}$$

and having the squared axes

$$\bar{\rho}_1^2 = r_0^2, \quad \bar{\eta}_1^2 = \left(\frac{r_0}{\sigma}\right)^2, \tag{313}$$

with

$$r_0^2 = \frac{\sigma B_1}{A^4|X(v)|^2}, \quad \sigma = \frac{1}{(1-\lambda^2/2)^2}. \tag{314}$$

Concerning $K_{\rm L} \to \pi^0\nu\bar{\nu}$, we may introduce—in analogy to (307)—the reduced branching ratio

$$B_2 \equiv \frac{1}{\kappa_{\rm L}}\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu}), \tag{315}$$

which is characterized by

$$\kappa_{\rm L} = \left[\frac{r_{K_{\rm L}}}{r_{K^+}}\frac{\tau_{K_{\rm L}}}{\tau_{K^+}}\right]\kappa_+ = 2.09 \times 10^{-10}, \tag{316}$$

where $r_{K_{\rm L}} = 0.944$ describes the isospin-breaking corrections that arise in relating $K_{\rm L} \to \pi^0\nu\bar{\nu}$ to $K^+ \to \pi^0 e^+\nu$. As discussed in detail in Ref. [37], the decay $K_{\rm L} \to \pi^0\nu\bar{\nu}$ is dominated in the SM by direct CP violation, and is completely governed by the short-distance loop diagrams with internal top-quark exchanges. Since the charm contribution can be fully neglected, the decay $K_{\rm L} \to \pi^0\nu\bar{\nu}$ is *even cleaner* than $K^+ \to \pi^+\nu\bar{\nu}$. In models with MFV, the reduced $K_{\rm L} \to \pi^0\nu\bar{\nu}$ branching ratio is given as follows:

$$B_2 = \left[\frac{\mathrm{Im}\lambda_t}{\lambda^5}|X(v)|\right]^2; \tag{317}$$

the SM corresponds to (49). If we now follow Ref. [45], but admit both signs of $X(v)$ and $S(v)$, we obtain

$$\bar{\rho} = 1 + \left[\frac{\pm\sqrt{\sigma(B_1-B_2)} + \mathrm{sgn}(X(v))P_c(\nu\bar{\nu})}{A^2|X(v)|}\right], \quad \bar{\eta} = \mathrm{sgn}(S(v))\frac{\sqrt{B_2}}{\sqrt{\sigma}A^2|X(v)|}. \tag{318}$$

The dependence on $|X(v)|$ cancels in the following quantity [126]:

$$r_s \equiv \frac{1-\overline{\rho}}{\overline{\eta}} = \mathrm{ctg}\beta = \mathrm{sgn}(S(v))\sqrt{\sigma}\left[\frac{\mp\sqrt{\sigma(B_1-B_2)} - \mathrm{sgn}(X(v))P_c(\nu\bar{\nu})}{\sqrt{B_2}}\right], \tag{319}$$

which allows the determination of $(\sin 2\beta)_{\pi\nu\bar{\nu}}$ in (48) through

$$\sin 2\beta = \frac{2r_s}{1+r_s^2}. \tag{320}$$

Note that (319) reduces to

$$r_s = \sqrt{\sigma}\left[\frac{\sqrt{\sigma(B_1-B_2)} - P_c(\nu\bar{\nu})}{\sqrt{B_2}}\right] \tag{321}$$

in the case of positive values of $S(v)$ and $X(v)$ [45]. Because of the relation in (186), it is actually more appropriate to consider the CP-violating observable $a_{\psi K_S}$ instead of $\sin 2\beta$. Consequently, we obtain a very interesting link between the mixing-induced CP violation in the 'golden' mode $B_d \to J/\psi K_S$ and the branching ratios of the rare $K \to \pi\nu\bar{\nu}$ decays.

Since $a_{\psi K_S}$ has already been measured with impressive accuracy and $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$ will be known rather accurately prior to the measurement of $\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})$, it is of particular interest to calculate $\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})$ as a function of $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$ for a given value of $a_{\psi K_S}$ [126]. To this end, it is useful to introduce the quantity

$$f(\beta) \equiv \mathrm{sgn}(S(v))\,\mathrm{ctg}\beta = \frac{1-\bar{\rho}}{|\bar{\eta}|}, \tag{322}$$

which can be determined *unambiguously* through

$$f(\beta) = \frac{1+\sqrt{1-a^2_{\psi K_S}}}{a_{\psi K_S}} = 2.279^{+0.235}_{-0.215}; \tag{323}$$

the numerical value corresponds to $a_{\psi K_S} = 0.736 \pm 0.049$. We then obtain the following expression:

$$B_1 = B_2 + \left[\frac{f(\beta)\sqrt{B_2} + \mathrm{sgn}(X(v))\sqrt{\sigma}P_c(\nu\bar{\nu})}{\sigma}\right]^2. \tag{324}$$

In comparison with (319), the advantage of (324) is the absence of the sign ambiguities due to $\mathrm{sgn}(S(v))$ and the $\mp$ in front of $\sqrt{\sigma(B_1 - B_2)}$. Consequently, for given values of $a_{\psi K_S}$ and $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$, only two values of $\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})$ are allowed for the full class of MFV models, independently of any new parameter present in these models. These two values of the $K_L \to \pi^0\nu\bar{\nu}$ branching ratio correspond to the two possible signs of $X(v)$. The measurement of $\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})$ will therefore either select one of these two possible values or will rule out all MFV models.

9.4 New Physics beyond minimal flavour violation: an example

As we have seen in Subsection 7.2.3, the pattern of the current B-factory data for the $B \to \pi K$ system suggests an enhancement of the corresponding EW penguin parameter q, and the presence of a CP-violating NP phase ϕ in the EW penguin sector, as summarized in (213). Since we encounter here CP-violating effects that are *not* associated with the CKM matrix, the corresponding NP does *not* belong to the category of MFV models considered above. In order to explore the implications for rare B and K decays, let us follow Refs. [48], [54], and consider a specific scenario, where the NP effects enter through enhanced Z^0 penguins, which are described by a short-distance function C.

The implications of enhanced Z^0 penguins with a large new complex phase for rare and CP-violating K and B decays were already discussed in Refs. [184]– [186], where model-independent analyses and studies within particular supersymmetric scenarios were presented. Here we determine the size of the enhancement of the Z^0-penguin function C and the magnitude of its complex phase through the $B \to \pi K$ data. As was pointed out in Ref. [146], a connection between rare decays and the $B \to \pi K$ system can be established by relating the EW penguin parameter q to the Z^0-penguin function C, which can be properly done with the help of a renormalization-group analysis. In the case of a complex EW penguin parameter, with a non-vanishing weak phase ϕ, we obtain the following relation [48], [54]:

$$C \equiv |C|e^{i\theta_C} = 2.35\;\bar{q}e^{i\phi} - 0.82, \quad \bar{q} = q\left[\frac{|V_{ub}/V_{cb}|}{0.086}\right]. \tag{325}$$

This quantity enters the short-distance functions X and Y, which govern the rare K, B decays with $\nu\bar{\nu}$ and $\mu^+\mu^-$ in the final states, respectively, in the linear combinations

$$X \equiv |X|e^{i\theta_X} = C + B^{\nu\bar{\nu}}, \quad Y \equiv |Y|e^{i\theta_Y} = C + B^{\mu^+\mu^-}, \tag{326}$$

where $B^{\nu\bar{\nu}}$ and $B^{\mu^+\mu^-}$ describe the box diagrams with $\nu\bar{\nu}$ and $\mu^+\mu^-$, respectively. If we evaluate, in the spirit of Refs. [146], [184], [185], these box-diagram contributions in the SM and use (325), we obtain

$$|X|e^{i\theta_X} = |C|e^{i\theta_C} + 0.73 \quad \text{and} \quad |Y|e^{i\theta_Y} = |C|e^{i\theta_C} + 0.18. \tag{327}$$

While the analysis described here does not rely on a particular model, concrete models with enhanced CP-violating Z^0-mediated FCNC couplings, generated either at the one-loop level or even at the tree level, were discussed in the literature (see, for instance, Refs. [9], [148], [184], [185], [186]). Let us also note that models with Z'-mediated FCNCs could be put in this class, provided their contributions can effectively be absorbed in the function C (for a recent analysis, see Ref. [187]).

If we now insert the numerical values in (213) into (327), we obtain a central value for $|Y|$ that violates the upper bound $|Y| \leq 2.2$ following from the BaBar and Belle data on $B \to X_s\mu^+\mu^-$ [188], and the upper bound on $\mathrm{BR}(K_{\rm L} \to \pi^0 e^+ e^-)$ of 2.8×10^{-10} from KTeV [189]. However, we may still encounter significant deviations from the SM. In order to illustrate this exciting feature, we consider only the subset of those values of (q, ϕ) in (213) that satisfy the constraint of $|Y| = 2.2$. If we then use (325) and (327), we obtain

$$\begin{array}{rclcrcl} |C| &=& 2.24 \pm 0.04, & \qquad & \theta_C &=& -(105 \pm 12)^\circ, \\ |X| &=& 2.17 \pm 0.12, & & \theta_X &=& -(87 \pm 12)^\circ, \\ |Y| &=& 2.2 \ (\mathrm{input}), & & \theta_Y &=& -(103 \pm 12)^\circ, \end{array} \tag{328}$$

which should be compared with the SM values $C(x_t) = 0.79$, $X(x_t) = 1.53$ and $Y(x_t) = 0.98$, corresponding to $m_t = 167$ GeV.

Going back now to the $B_q \to \mu^+\mu^-$ decays, we find

$$\frac{\mathrm{BR}(B_s \to \mu^+\mu^-)}{\mathrm{BR}(B_s \to \mu^+\mu^-)_{\rm SM}} = \frac{\mathrm{BR}(B_d \to \mu^+\mu^-)}{\mathrm{BR}(B_d \to \mu^+\mu^-)_{\rm SM}} = \left|\frac{Y}{Y_{\rm SM}}\right|^2 \approx 5.0. \tag{329}$$

This significant enhancement corresponds to the branching ratios

$$\mathrm{BR}(B_s \to \mu^+\mu^-) \approx 17 \times 10^{-9}, \quad \mathrm{BR}(B_d \to \mu^+\mu^-) \approx 5 \times 10^{-10}, \tag{330}$$

which are still well below the experimental bounds summarized in (288).

As far as the $K \to \pi\nu\bar{\nu}$ decays are concerned, this NP analysis implies

$$\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu}) = (7.5 \pm 2.1) \times 10^{-11}, \quad \mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu}) = (3.1 \pm 1.0) \times 10^{-10}, \tag{331}$$

which should be compared with the SM predictions in (49). We observe that the impact of NP on the $K^+ \to \pi^+\nu\bar{\nu}$ branching ratio would be small, whereas $\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu})$ would be dramatically enhanced. If we introduce

$$\beta_X \equiv \beta - \beta_s - \theta_X \quad \text{with} \quad \beta_s \equiv -\delta\gamma = -\lambda^2\eta, \tag{332}$$

we see that this exciting pattern is dominantly the consequence of $\beta_X \approx 111^\circ$, as

$$\frac{\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu})}{\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu})_{\rm SM}} = \left|\frac{X}{X_{\rm SM}}\right|^2 \left[\frac{\sin\beta_X}{\sin(\beta - \beta_s)}\right]^2 \tag{333}$$

and

$$\frac{\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu})}{\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})} \approx 4.4 \times (\sin\beta_X)^2 \approx 4.2 \pm 0.2. \tag{334}$$

It is interesting to note that $\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu})$ is very close to its model-independent upper bound [190]:

$$\mathrm{BR}(K_{\rm L} \to \pi^0\nu\bar{\nu}) \leq 4.4 \times \mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu}). \tag{335}$$

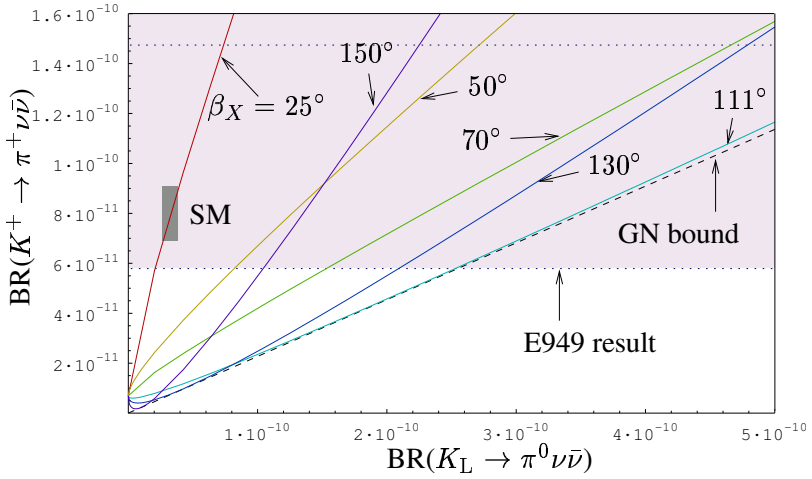


Fig. 27: $\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$ as a function of $\mathrm{BR}(K_\mathrm{L} \to \pi^0\nu\bar{\nu})$ for various values of β_X. The dotted horizontal lines indicate the experimental range (50) and the grey area the SM prediction. We also show the bound in (335).

A spectacular implication of these findings is a strong violation of the relation in (48). Indeed,

$$(\sin 2\beta)_{\pi\nu\bar{\nu}} = \sin 2\beta_X = -(0.69^{+0.23}_{-0.41}), \tag{336}$$

in striking disagreement with $(\sin 2\beta)_{\psi K_\mathrm{S}} = 0.736 \pm 0.049$. In Fig. 27, we plot—in the spirit of Ref. [126]—$\mathrm{BR}(K^+ \to \pi^+\nu\bar{\nu})$ as a function of $\mathrm{BR}(K_\mathrm{L} \to \pi^0\nu\bar{\nu})$ for fixed values of β_X. As this plot is independent of $|X|$, it offers a direct measurement of the phase β_X. The first line on the left represents the MFV models with $\beta_X = \beta - \beta_s$, whereas the first line on the right corresponds to the model-independent Grossman–Nir bound given in (335). The central value $\beta_X = 111°$ found in Refs. [48], [54] is very close to this bound. As can be seen in Fig. 27, the measured $K \to \pi\nu\bar{\nu}$ branching ratios allow us to determine β_X up to discrete ambiguities, which can be resolved by considering other rare decays simultaneously. The corresponding plot for different values of β_X that are close to β can be found in Ref. [126].

In addition to the significant and—in the case of $K_\mathrm{L} \to \pi^0\nu\bar{\nu}$ and $(\sin 2\beta)_{\pi\nu\bar{\nu}}$—even spectacular NP effects discussed above, there are further interesting implications of this scenario [48], [54]:

- The branching ratio
$$\mathrm{BR}(K_\mathrm{L} \to \pi^0 e^+ e^-) = (7.8 \pm 1.6) \times 10^{-11} \tag{337}$$
is significantly enhanced and governed by direct CP violation. On the other hand, the SM result $(3.2^{+1.2}_{-0.8}) \times 10^{-11}$ [191] is dominated by indirect CP violation. In a very recent analysis [192], the same NP scenario was considered as well, addressing also the decay $K_\mathrm{L} \to \pi^0\mu^+\mu^-$.
- The integrated forward–backward CP asymmetry for $B_d \to K^*\mu^+\mu^-$ [186], which is given by
$$A_\mathrm{FB}^\mathrm{CP} = (0.03 \pm 0.01) \times \tan\theta_Y, \tag{338}$$
can be very large in view of $\theta_Y \approx -100°$. The corresponding NP effects for the lepton polarization asymmetries of $B \to X_s\ell^+\ell^-$ decays were recently studied in Ref. [193].
- The $B \to X_{s,d}\nu\bar{\nu}$ branching ratios are enhanced by a factor of 2 with respect to the SM.
- Enhanced Z^0 penguins may also play an important role in $\mathrm{Re}(\varepsilon'/\varepsilon)$ [184]. As far as the enhancement of $|C|$ and its large negative phase suggested by the $B \to \pi K$ analysis are concerned, the consistency with (47) requires a significant enhancement of the hadronic matrix element of the relevant QCD penguin operator with respect to that of the relevant EW penguin operator. The corresponding large hadronic uncertainties leave sufficient room for such effects.

– It is also interesting to explore the implications for $B_d \to J/\psi K_S$ and $B_d \to \phi K_S$. As far as the former channel is concerned, the NP corrections to the determination of $\sin 2\beta$ from the mixing-induced $B_d \to J/\psi K_S$ CP asymmetry are at the 0.05 level, corresponding to a shift of β by at most $\pm 2^\circ$. Such small effects are still beyond the current experimental and theoretical accuracy, but could be reinvestigated in the LHC era. Concerning the decay $B_d \to \phi K_S$, large hadronic uncertainties preclude a precise prediction. However, if we assume that the sign of the cosine of a strong phase agrees with factorization, we find

$$\underbrace{(\sin 2\beta)_{\phi K_S}}_{-\mathcal{A}_{CP}^{mix}(B_d \to \phi K_S)} > \underbrace{(\sin 2\beta)_{\psi K_S} = 0.736 \pm 0.049}_{-\mathcal{A}_{CP}^{mix}(B_d \to J/\psi K_S)}, \tag{339}$$

where $(\sin 2\beta)_{\phi K_S} \sim 1$ may well be possible. This pattern is qualitatively different from the present B-factory data summarized in (179), which are, however, not yet conclusive. In particular, we could easily accommodate a value of $(\sin 2\beta)_{\phi K_S}$ of the same magnitude as the central value found by Belle but of *opposite* sign. On the other hand, a future confirmation of the pattern in (339) would be another signal of enhanced CP-violating Z^0 penguins at work.

If future, more accurate $B \to \pi\pi, \pi K$ data will not significantly modify the currently observed patterns in these decays discussed in Subsections 6.2.2 and 7.2.3, the scenario of enhanced Z^0 penguins with a large CP-violating NP phase ϕ will remain an attractive possibility for physics beyond the SM. It will then be very interesting to confront the corresponding predictions for the rare B and K decays discussed above with experiment.

10 CONCLUSIONS

The field of flavour physics and CP violation is very rich and represents an exciting topic for theoretical and experimental research. In these lectures, we have put our focus on the B-meson system, which provides a particularly fertile testing ground for the SM picture of flavour physics, where CP violation can be accommodated by means of the KM mechanism through a single phase in the parametrization of the quark-mixing matrix. The corresponding UT represents one of the central targets of the B factories, which govern the current experimental stage of quark-flavour physics, run II of the Tevatron, and of the LHCb and BTeV experiments, which will join these efforts in the not too distant future.

In 1964, the observation of indirect CP violation, which originates from the fact that the mass eigenstates of the neutral kaon system are not eigenstates of the CP operator, came as a big surprise. After tremendous efforts, also direct CP violation could be established in neutral K decays in 1999 by the NA48 and KTeV collaborations. Unfortunately, the calculations of the corresponding observable $\mathrm{Re}(\varepsilon'/\varepsilon)$, which is governed by the competition between QCD and EW penguins, suffer from large theoretical uncertainties. Consequently, unless better techniques to deal with the relevant hadronic matrix elements become available, $\mathrm{Re}(\varepsilon'/\varepsilon)$ unfortunately does not provide a stringent test of the SM, although the SM analyses give results of the same order of magnitude as the experimental value. From the theoretical point of view, the rare decays $K^+ \to \pi^+ \nu\bar{\nu}$ and $K_L \to \pi^0 \nu\bar{\nu}$ are much more promising. On the other hand, these decays exhibit extremely tiny branching ratios at the 10^{-10} and 10^{-11} levels in the SM, respectively, and are extremely challenging from the experimental point of view. Nevertheless, three events for $K^+ \to \pi^+ \nu\bar{\nu}$ were already observed at BNL.

Concerning the decays of B mesons, we distinguish between leptonic, semileptonic and non-leptonic transitions. The former exhibit the simplest structure and would be interesting to measure the non-perturbative decay constants f_B, but suffer from tiny branching ratios. The semileptonic B decays are more complicated than the leptonic ones. However, applications of the HQET and heavy-quark expansions allow us to determine $|V_{cb}|$ and $|V_{ub}|$, which are important ingredients for theoretical predictions and the analysis of the UT in the $\bar{\rho}$–$\bar{\eta}$ plane. Finally, the non-leptonic decays are the most complicated transitions, as far as the impact of strong interactions is concerned. In order to deal with them theoretically, low-energy effective Hamiltonians are used, which consist of perturbatively calculable Wilson

coefficients and local four-quark operators. The former encode the whole short-distance dynamics of the decay class at hand, whereas the long-distance contributions of a specific channel show up as the corresponding hadronic matrix elements of the four-quark operators. The same formalism applies of course also to non-leptonic kaon decays and is at the basis of the calculations of $\mathrm{Re}(\varepsilon'/\varepsilon)$. The non-leptonic B decays play the key role for the exploration of CP violation, since non-vanishing CP asymmetries may be induced by interference effects in such transitions. In general, the theoretical interpretation of such CP asymmetries is affected by large hadronic uncertainties, in analogy to $\mathrm{Re}(\varepsilon'/\varepsilon)$. However, the B-meson system provides tools to deal with these uncertainties: there are fortunate cases, where relations between various decay amplitudes allow us to *eliminate* the—essentially unknown—hadronic matrix elements, and we may exploit mixing-induced CP asymmetries, where the hadronic matrix elements *cancel* if the decay is governed by a single CKM amplitude. The latter observables can also be nicely combined with amplitude relations. Following these lines, we may also determine—in addition to the angles of the UT—certain hadronic parameters, which can then be compared with the corresponding theoretical calculations, where also a lot of progress could be made over recent years.

Thanks to the efforts of the BaBar and Belle collaborations, CP violation could be established in the B-meson system in 2001, with the help of the 'golden' mode $B_d \to J/\psi K_{\rm S}$, thereby opening a new era in the exploration of this phenomenon. The current experimental status of the mixing-induced CP asymmetry of this (and similar) channel(s) implies $\sin 2\beta = 0.736 \pm 0.049$, in impressive accordance with the indirect value following from the CKM fits of the UT in the $\bar{\rho}$–$\bar{\eta}$ plane. The physics potential of the B factories goes far beyond the famous $B_d \to J/\psi K_{\rm S}$ decay, allowing us now to confront many more strategies to explore CP violation with data. Here the main goal is to overconstrain the UT as much as possible, thereby performing a stringent test of the KM mechanism of CP violation. Important B-factory benchmark modes to complement the $B \to J/\psi K$ system are given by $B \to \pi\pi$ and $B \to \phi K$ decays, and exciting data on these channels are already available. The pattern of the $B \to \pi\pi$ data favours large non-factorizable effects, and the analyses of CP violation in $B_d \to \pi^+\pi^-$ point towards large direct and mixing-induced CP asymmetries, which can be interpreted in terms of $\gamma \sim 65^\circ$, in accordance with the CKM fits. Although the BaBar and Belle measurements of these asymmetries are not yet in full accordance, they already moved towards each other and it seems plausible that they will meet close to the current averages. On the other hand, the Belle measurement of the mixing-induced CP asymmetry of $B_d \to \phi K_{\rm S}$ raises the exciting possibility of having large NP effects in the $\bar{b} \to \bar{s}s\bar{s}$ quark-level processes. However, the corresponding BaBar analysis is consistent with the SM, so that we cannot yet draw firm conclusions. Let us hope that this unsatisfactory experimental situation will be clarified soon.

As far as the exploration of CP violation with the help of amplitude relations is concerned, we distinguish between exact and flavour-symmetry relations. The prototype of the former is provided by $B^\pm \to K^\pm D$ decays, whereas $B_c^\pm \to D_s^\pm D$ transitions offer the ideal theoretical realization of the corresponding triangle strategy to determine the angle γ of the UT. An important example for the application of flavour-symmetry relations is given by $B \to \pi K$ decays. Here the corresponding B-factory data point again to a puzzling pattern, which may be due to the presence of enhanced EW penguins with a large CP-violating NP phase. Although BaBar, Belle and CLEO indicate separately the corresponding '$B \to \pi K$ puzzle', it is still too early for definite conclusions. This kind of NP would yield striking effects in various rare B and K decays, of which an enhancement of the $K_{\rm L} \to \pi^0\nu\bar{\nu}$ branching ratio by one order of magnitude and a negative value of $(\sin 2\beta)_{\pi\nu\bar{\nu}}$ would be the most spectacular ones.

Another key element for the testing of the SM description of CP violation is the B_s-meson system, which is not accessible at the e^+e^- B factories operating at the $\Upsilon(4S)$ resonance, BaBar and Belle, but can be studied nicely at hadron collider experiments. Interesting results on B_s physics are soon expected from run II of the Tevatron, where B_s^0–$\bar{B}_s^0$ mixing should be discovered, which is an important ingredient for the CKM fits of the UT. The most prominent B_s decays include $B_s \to J/\psi\phi$, which is a powerful probe for NP contributions to B_s^0–$\bar{B}_s^0$ mixing manifesting themselves through a sizeable value of ϕ_s; $B_s \to K^+K^-$, which can be combined with $B_d \to \pi^+\pi^-$ through the U-spin flavour symmetry to

determine γ; and $B_s \to D_s^{(*)\pm} K^{\mp}$ modes, which allow clean determinations of $\phi_s + \gamma$ and can be combined in a variety of ways with their $B_d \to D^{(*)\pm}\pi^{\mp}$ counterparts, offering advantages from the practical point of view. Although the Tevatron will provide first insights into these decays, they can only be fully exploited at the experiments of the LHC era, in particular LHCb and BTeV.

Finally, it should be emphasized again that it is crucial to complement the studies of CP violation with measurements of rare B and K decays, which are sensitive probes for NP. Moreover, it is important to keep also an eye on the D-meson system, which exhibits tiny mixing and CP-violating effects in the SM [121], as well as on various other interesting aspects of flavour physics, such as flavour-violating charged-lepton decays (for a very recent study, see Ref. [194]), which we could not cover in these lectures.

In this decade, the successful exploration of flavour physics and CP violation will certainly be continued, thereby leading to many further exciting results and valuable new insights. Let us hope that eventually also several 'surprises' can be established, shedding light on the physics beyond the SM!

ACKNOWLEDGEMENTS

I would like to thank the students for their interest in my lectures, the discussion leaders for their efforts to complement them in the discussion sessions, and the local organizers for hosting this exciting school in Armenia. I am also grateful to my collaborators for the fun we had working on many of the topics addressed in these lectures.

REFERENCES

[1] J.H. Christenson, J.W. Cronin, V.L. Fitch and R. Turlay, Phys. Rev. Lett. **13** (1964) 138.

[2] V. Fanti *et al.* [NA48 Collaboration], Phys. Lett. **B465** (1999) 335.

[3] A. Alavi-Harati *et al.* [KTeV Collaboration], Phys. Rev. Lett. **83** (1999) 22.

[4] B. Aubert *et al.* [BaBar Collaboration], Phys. Rev. Lett. **87** (2001) 091801.

[5] K. Abe *et al.* [Belle Collaboration], Phys. Rev. Lett. **87** (2001) 091802.

[6] I. Aitchison, lectures given at this school.

[7] G. Gabadadze, lectures given at this school.

[8] Y. Grossman, Int. J. Mod. Phys. **A19** (2004) 907;
J.R. Ellis, CERN-TH/2002-339 [hep-ph/0211322], talk given at 1st International Workshop on Frontier Science: Charm, Beauty, and CP, Frascati, Rome, Italy, 6–11 October 2002;
A. Masiero and O. Vives, Annu. Rev. Nucl. Part. Sci. **51** (2001) 161;
L. Wolfenstein, Phys. Rev. **D57** (1998) 6857;
M. Gronau and D. London, Phys. Rev. **D55** (1997) 2845;
Y. Nir and H.R. Quinn, Annu. Rev. Nucl. Part. Sci. **42** (1992) 211.

[9] Y. Grossman, Y. Nir and R. Rattazzi, Adv. Ser. Direct. High Energy Phys. **15** (1998) 755.

[10] S. Petkov, lectures given at this school.

[11] G. Altarelli and F. Feruglio, hep-ph/0306265.

[12] A. De Rújula, M.B. Gavela and P. Hernandez, *Nucl. Phys.* **B547** (1999) 21;
K. Dick, M. Freund, M. Lindner and A. Romanino, *Nucl. Phys.* **B562** (1999) 29;
P. Huber, M. Lindner, M. Rolinec, T. Schwetz and W. Winter, hep-ph/0403068.

[13] I. Tkachev, lectures given at this school.

[14] For a recent discussion, see also W. Buchmüller, DESY-03-068 [hep-ph/0306047].

[15] A.D. Sakharov, JETP Lett. **5** (1967) 24.

[16] V.A. Rubakov, M.E. Shaposhnikov, Usp. Fiz. Nauk **166** (1996) 493; Phys. Usp. **39** (1996) 461; A. Riotto and M. Trodden, Annu. Rev. Nucl. Part. Sci. **49** (1999) 35.

[17] *The BaBar Physics Book*, eds. P. Harrison and H.R. Quinn, SLAC-R-504 (1998).

[18] K. Anikeev *et al.*, FERMILAB-Pub-01/197 [hep-ph/0201071].

[19] P. Ball *et al.*, CERN-TH/2000-101 [hep-ph/0003238], in CERN Report on *Standard Model physics (and more) at the LHC* (CERN, Geneva, 2000), p. 305.

[20] G. Branco, L. Lavoura and J. Silva, *CP Violation* (Oxford Science Publications, Clarendon Press, Oxford, 1999); I.I. Bigi and A.I. Sanda, *CP Violation* (Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, Cambridge, 2000).

[21] A.J. Buras and R. Fleischer, Adv. Ser. Direct. High Energy Phys. **15** (1998) 65.

[22] A.J. Buras, hep-ph/9806471, lectures given at Summer School on Theoretical Physics: Probing the Standard Model of Particle Interactions, Les Houches, France, 28 July – 5 September 1997.

[23] Y. Nir, hep-ph/9911321, lectures given at 27th SLAC Summer Institute on Particle Physics: CP Violation in and Beyond the Standard Model (SSI 99), Stanford, CA, 7–16 July 1999.

[24] J. Rosner, hep-ph/0011355, lectures given at Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2000): Flavor Physics for the Millenium, Boulder, CO, 4–30 June 2000.

[25] Z. Ligeti, hep-ph/0302031, lectures given at 30th SLAC Summer Institute on Particle Physics: Secrets of the B Meson (SSI 2002), Stanford, CA, 5–16 August 2002.

[26] R. Fleischer, Phys. Rep. **370** (2002) 531.

[27] S.L. Glashow, Nucl. Phys. **22** (1961) 579;
S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264;
A. Salam, in *Elementary Particle Theory*, ed. N. Svartholm (Almqvist and Wiksell, Stockholm, 1968).

[28] N. Cabibbo, Phys. Rev. Lett. **10** (1963) 531.

[29] M. Kobayashi and T. Maskawa, Progr. Theor. Phys. **49** (1973) 652.

[30] S.L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. **D2** (1970) 1285.

[31] K. Hagiwara *et al.* [Particle Data Group], Phys. Rev. **D66** (2002) 010001.

[32] H. Fritzsch and Z.-Z. Xing, Phys. Lett. **B413** (1997) 396.

[33] C. Jarlskog, Phys. Rev. Lett. **55** (1985) 1039; Z. Phys. **C29** (1985) 491.

[34] J. Bernabeu, G. Branco and M. Gronau, Phys. Lett. **B169** (1986) 243.

[35] L. Wolfenstein, Phys. Rev. Lett. **51** (1983) 1945.

[36] A.J. Buras, M.E. Lautenbacher and G. Ostermaier, Phys. Rev. **D50** (1994) 3433.

[37] A.J. Buras, hep-ph/0101336, lectures given at Erice International School of Subnuclear Physics: Theory and Experiment Heading for New Physics, Erice, Italy, 27 August – 5 September 2000.

[38] R. Aleksan, B. Kayser and D. London, Phys. Rev. Lett. **73** (1994) 18.

[39] C. Jarlskog and R. Stora, Phys. Lett. **B208** (1988) 268.

[40] L.L. Chau and W.-Y. Keung, Phys. Rev. Lett. **53** (1984) 1802.

[41] M. Battaglia *et al.*, CERN 2003-002-corr [hep-ph/0304132].

[42] J.R. Batley *et al.* [NA48 Collaboration], Phys. Lett. **B544** (2002) 97.

[43] A. Alavi-Harati *et al.* [KTeV Collaboration], Phys. Rev. **D67** (2003) 012005.

[44] A.J. Buras and M. Jamin, JHEP **0401** (2004) 048.

[45] G. Buchalla and A.J. Buras, Phys. Lett. **B333** (1994) 221; Phys. Rev. **D54** (1996) 6782.

[46] G. D'Ambrosio and G. Isidori, Phys. Lett. **B530** (2002) 108.

[47] R. Fleischer, G. Isidori and J. Matias, JHEP **0305** (2003) 053.

[48] A.J. Buras, R. Fleischer, S. Recksiegel, F. Schwab, CERN-PH-TH/2004-020 [hep-ph/0402112].

[49] G. Isidori, eConf **C0304052** (2003) WG304 [hep-ph/0307014].

[50] S.H. Kettell, L.G. Landsberg and H.H. Nguyen, FERMILAB-FN-0727 [hep-ph/0212321].

[51] V.V. Anisimovsky *et al.* [E949 Collaboration], BNL/72164-2004-JA [hep-ex/0403036].

[52] S. Adler *et al.* [E787 Collaboration], Phys. Rev. Lett. **88** (2002) 041803; BNL-72163-2004-JA [hep-ex/0403034].

[53] A. Alavi-Harati *et al.* [E799-II/KTeV Collaboration], Phys. Rev. **D61** (2000) 072006.

[54] A.J. Buras, R. Fleischer, S. Recksiegel and F. Schwab, *Phys. Rev. Lett.* **92** (2004) 101804.

[55] D.E. Jaffe, hep-ex/0311053.

[56] M. Lüscher, Annales Henri Poincaré **4** (2003) S197 [hep-ph/0211220].

[57] A. Khodjamirian, lectures given at this school [hep-ph/0403145].

[58] F. De Fazio, hep-ph/0010007.

[59] D.G. Cassel, eConf **C0304052** (2003) WG501 [hep-ex/0307038].

[60] N. Isgur and M.B. Wise, Phys. Lett. **B232** (1989) 113 and **B237** (1990) 527.

[61] M. Neubert, Phys. Rep. **245** (1994) 259.

[62] M. Neubert, Phys. Lett. **B264** (1991) 455.

[63] M.E. Luke, Phys. Lett. **B252** (1990) 447.

[64] F.J. Gilman and M.B. Wise, Phys. Rev. **D20** (1979) 2392;
G. Altarelli, G. Curci, G. Martinelli and S. Petrarca, Phys. Lett. **B99** (1981) 141;
A.J. Buras and P.H. Weisz, Nucl. Phys. **B333** (1990) 66.

[65] G. Buchalla, A.J. Buras and M.E. Lautenbacher, Rev. Mod. Phys. **68** (1996) 1125.

[66] M. Bander, D. Silverman and A. Soni, Phys. Rev. Lett. **43** (1979) 242.

[67] R. Fleischer, Z. Phys. **C58** (1993) 483.

[68] A.J. Buras and R. Fleischer, Phys. Lett. **B341** (1995) 379.

[69] M. Ciuchini, E. Franco, G. Martinelli, M. Pierini and L. Silvestrini, Phys. Lett. **B515** (2001) 33;
C. Isola, M. Ladisa, G. Nardulli, T.N. Pham and P. Santorelli, Phys. Rev. **D65** (2002) 094005;
C.W. Bauer, D. Pirjol, I.Z. Rothstein and I.W. Stewart, MIT-CTP-3469 [hep-ph/0401188].

[70] R. Fleischer, Z. Phys. **C62** (1994) 81; Phys. Lett. **B321** (1994) 259 and **B332** (1994) 419.

[71] R. Fleischer, Int. J. Mod. Phys. **A12** (1997) 2459.

[72] N.G. Deshpande and X.-G. He, Phys. Rev. Lett. **74** (1995) 26 [E: ibid., p. 4099];
M. Gronau, O.F. Hernandez, D. London and J.L. Rosner, Phys. Rev. **D52** (1995) 6374.

[73] M. Neubert, B. Stech, Adv. Ser. Direct. High Energy Phys. **15** (1998) 294, and references therein.

[74] A.J. Buras and J.-M. Gérard, Nucl. Phys. **B264** (1986) 371;
A.J. Buras, J.-M. Gérard and R. Rückl, Nucl. Phys. **B268** (1986) 16.

[75] M. Beneke, G. Buchalla, M. Neubert and C. Sachrajda, Phys. Rev. Lett. **83** (1999) 1914; Nucl. Phys. **B591** (2000) 313; Nucl. Phys. **B606** (2001) 245.

[76] J.D. Bjorken, Nucl. Phys. (Proc. Suppl.) **B11** (1989) 325;
M. Dugan and B. Grinstein, Phys. Lett. **B255** (1991) 583;
H.D. Politzer and M.B. Wise, Phys. Lett. **B257** (1991) 399.

[77] H.-n. Li and H.L. Yu, Phys. Rev. **D53** (1996) 2480;
Y.Y. Keum, H.-n. Li and A.I. Sanda, Phys. Lett. **B504** (2001) 6;
Y.Y. Keum and H.-n. Li, Phys. Rev. **D63** (2001) 074006;
Y.Y. Keum and A.I. Sanda, eConf **C0304052** (2003) WG420 [hep-ph/0306004].

[78] C.W. Bauer, D. Pirjol and I.W. Stewart, Phys. Rev. Lett. **87** (2001) 201806;
C.W. Bauer, B. Grinstein, D. Pirjol and I.W. Stewart, Phys. Rev. **D67** (2003) 014010.

[79] A. Khodjamirian, Nucl. Phys. **B605** (2001) 558;

A. Khodjamirian, T. Mannel and B. Melic, Phys. Lett. **B571** (2003) 75.

[80] A.J. Buras, M. Jamin and P.H. Weisz, Nucl. Phys. **B347** (1990) 491;
J. Urban, F. Krauss, U. Jentschura and G. Soff, Nucl. Phys. **B523** (1998) 40.

[81] T. Inami and C.S. Lim, Prog. Theor. Phys. **65** (1981) 297 [E: ibid., p. 1772].

[82] A.J. Buras, hep-ph/0307203, lectures given at 41st International University School of Theoretical Physics: Flavour Physics (IUTP 41), Schladming, Styria, Austria, 22–28 February 2003.

[83] S. Laplace, Z. Ligeti, Y. Nir and G. Perez, Phys. Rev. **D65** (2002) 094040.

[84] M. Beneke, G. Buchalla, A. Lenz and U. Nierste, Phys. Lett. **B576** (2003) 173;
M. Ciuchini, E. Franco, V. Lubicz, F. Mescia and C. Tarantino, JHEP **0308** (2003) 031.

[85] B Oscillations Working Group: `http://lepbosc.web.cern.ch/LEPBOSC/`.

[86] Heavy Flavour Averaging Group: `http://www.slac.stanford.edu/xorg/hfag/`.

[87] R. Fleischer, Eur. Phys. J. **C10** (1999) 299.

[88] A.B. Carter and A.I. Sanda, Phys. Rev. Lett. **45** (1980) 952, Phys. Rev. **D23** (1981) 1567;
I.I. Bigi and A.I. Sanda, Nucl. Phys. **B193** (1981) 85.

[89] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **89** (2002) 201802.

[90] K. Abe *et al.* [Belle Collaboration], BELLE-CONF-0353 [hep-ex/0308036].

[91] R. Fleischer and T. Mannel, Phys. Lett. **B506** (2001) 311.

[92] H. Boos, T. Mannel and J. Reuter, SI-HEP-2004-04 [hep-ph/0403085].

[93] Ya.I. Azimov, V.L. Rappoport and V.V. Sarantsev, Z. Phys. **A356** (1997) 437;
Y. Grossman and H.R. Quinn, Phys. Rev. **D56** (1997) 7259;
J. Charles *et al.*, Phys. Lett. **B425** (1998) 375 [E: **B433** (1998) 441];
B. Kayser and D. London, Phys. Rev. **D61** (2000) 116012;
H.R. Quinn, T. Schietinger, J.P. Silva and A.E. Snyder, Phys. Rev. Lett. **85** (2000) 5284.

[94] A.S. Dighe, I. Dunietz and R. Fleischer, Phys. Lett. **B433** (1998) 147.

[95] I. Dunietz, R. Fleischer and U. Nierste, Phys. Rev. **D63** (2001) 114015.

[96] R. Fleischer, Phys. Lett. **B459** (1999) 306.

[97] M. Gronau and D. London, Phys. Rev. Lett. **65** (1990) 3381.

[98] J.P. Silva and L. Wolfenstein, Phys. Rev. **D49** (1994) 1151.

[99] R. Fleischer and T. Mannel, Phys. Lett. **B397** (1997) 269.

[100] Y. Grossman and H.R. Quinn, Phys. Rev. **D58** (1998) 017504.

[101] J. Charles, Phys. Rev. **D59** (1999) 054007.

[102] M. Gronau, D. London, N. Sinha and R. Sinha, Phys. Lett. **B514** (2001) 315

[103] H. Jawahery, talk given at Lepton–Photon 2003, Fermilab, Batavia, IL, 11–16 August 2003, `http://conferences.fnal.gov/lp2003/`.

[104] K. Abe *et al.* [Belle Collaboration], Phys. Rev. **D68** (2003) 012001.

[105] R. Fleischer and J. Matias, Phys. Rev. **D66** (2002) 054009.

[106] B. Aubert *et al.* [BaBar Collaboration], Phys. Rev. Lett. **91** (2003) 241801.

[107] K. Abe *et al.* [Belle Collaboration], Phys. Rev. Lett. **91** (2003) 261801.

[108] M. Beneke and M. Neubert, Nucl. Phys. **B675** (2003) 333.

[109] A. Ali, E. Lunghi and A.Y. Parkhomenko, DESY-04-036 [hep-ph/0403275].

[110] C.W. Chiang, M. Gronau, J.L. Rosner and D.A. Suprun, MADPH-04-1372 [hep-ph/0404073].

[111] G. Buchalla and A.S. Safir, LMU-25-03 [hep-ph/0310218];
F.J. Botella and J.P. Silva, hep-ph/0312337.

[112] D. London and R.D. Peccei, Phys. Lett. **B223** (1989) 257;
N.G. Deshpande and J. Trampetic, Phys. Rev. **D41** (1990) 895 and 2926;

J.-M. Gérard and W.-S. Hou, Phys. Rev. **D43** (1991) 2909; Phys. Lett. **B253** (1991) 478.
[113] N.G. Deshpande and X.-G. He, Phys. Lett. **B336** (1994) 471.
[114] Y. Grossman and M.P. Worah, Phys. Lett. **B395** (1997) 241.
[115] D. London and A. Soni, Phys. Lett. **B407** (1997) 61.
[116] R. Fleischer and T. Mannel, Phys. Lett. **B511** (2001) 240.
[117] T. Browder, hep-ex/0312024, talk given at Lepton–Photon 2003, Fermilab, Batavia, IL, 11–16 August 2003, `http://conferences.fnal.gov/lp2003/`.
[118] B. Aubert *et al.* [BaBar Collaboration], BABAR-PUB-04-004 [hep-ex/0403026].
[119] K. Abe *et al.* [Belle Collaboration], Phys. Rev. Lett. **91** (2003) 261602.
[120] G. Hiller, Phys. Rev. **D66** (2002) 071502;
A. Datta, Phys. Rev. **D66** (2002) 071702;
M. Raidal, Phys. Rev. Lett. **89** (2002) 231803;
B. Dutta, C.S. Kim and S. Oh, Phys. Rev. Lett. **90** (2003) 011801;
C.W. Chiang and J.L. Rosner, Phys. Rev. **D68** (2003) 014007;
C.K. Chua, W.S. Hou and M. Nagashima, hep-ph/0308298.
[121] A.A. Petrov, WSU-HEP-0314 [hep-ph/0311371].
[122] A.J. Buras, Acta Phys. Polon. **B34** (2003) 5615.
[123] A.J. Buras, P. Gambino, M. Gorbahn, S. Jäger and L. Silvestrini, Phys. Lett. **B500** (2001) 161.
[124] G. D'Ambrosio, G.F. Giudice, G. Isidori and A. Strumia, Nucl. Phys. **B645** (2002) 155.
[125] C. Bobeth, T. Ewerth, F. Krüger and J. Urban, Phys. Rev. **D66** (2002) 074021.
[126] A.J. Buras and R. Fleischer, Phys. Rev. **D64** (2001) 115010.
[127] A.J. Buras and R. Buras, Phys. Lett. **B501** (2001) 223.
[128] M. Gronau and D. Wyler, Phys. Lett. **B265** (1991) 172.
[129] D. Atwood, I. Dunietz, A. Soni, Phys. Rev. Lett. **78** (1997) 3257; Phys. Rev. **D63** (2001) 036005.
[130] F. Abe *et al.* [CDF Collaboration], Phys. Rev. Lett. **81** (1998) 2432.
[131] M. Masetti, Phys. Lett. **B286** (1992) 160.
[132] R. Fleischer and D. Wyler, Phys. Rev. **D62** (2000) 057503.
[133] M.A. Ivanov, J.G. Körner and O.N. Pakhomova, Phys. Lett. **B555** (2003) 189.
[134] R. Fleischer, Phys. Lett. **B365** (1996) 399.
[135] R. Fleischer and T. Mannel, Phys. Rev. **D57** (1998) 2752.
[136] M. Gronau and J.L. Rosner, Phys. Rev. **D57** (1998) 6843.
[137] R. Fleischer, Eur. Phys. J. **C6** (1999) 451.
[138] M. Neubert and J.L. Rosner, Phys. Lett. **B441** (1998) 403; Phys. Rev. Lett. **81** (1998) 5076.
[139] M. Neubert, JHEP **9902** (1999) 014.
[140] A.J. Buras and R. Fleischer, Eur. Phys. J. **C11** (1999) 93.
[141] A.J. Buras and R. Fleischer, Eur. Phys. J. **C16** (2000) 97.
[142] M. Gronau, J.L. Rosner and D. London, Phys. Rev. Lett. **73** (1994) 21.
[143] R. Fleischer and J. Matias, Phys. Rev. **D61** (2000) 074004.
[144] T. Yoshikawa, Phys. Rev. **D68** (2003) 054023.
[145] M. Gronau and J.L. Rosner, Phys. Lett. **B572** (2003) 43.
[146] A.J. Buras, R. Fleischer, S. Recksiegel and F. Schwab, Eur. Phys. J. **C32** (2003) 45.
[147] R. Fleischer and T. Mannel, TTP-97-22 [hep-ph/9706261].
[148] Y. Grossman, M. Neubert and A.L. Kagan, JHEP **9910** (1999) 029.
[149] B. Aubert *et al.* [BaBar Collaboration], BABAR-PUB-04-005 [hep-ex/0403001].

[150] A.A. Penin and M. Steinhauser, Phys. Rev. **D65** (2002) 054006;
M. Jamin and B.O. Lange, Phys. Rev. **D65** (2002) 056005;
K. Hagiwara, S. Narison and D. Nomura, Phys. Lett. **B540** (2002) 233.

[151] I. Dunietz, Phys. Rev. **D52** (1995) 3048.

[152] R. Fleischer and I. Dunietz, Phys. Rev. **D55** (1997) 259.

[153] R. Fleischer and I. Dunietz, Phys. Lett. **B387** (1996) 361.

[154] A.S. Dighe, I. Dunietz, H.J. Lipkin and J.L. Rosner, Phys. Lett. **B369** (1996) 144.

[155] A.S. Dighe, I. Dunietz and R. Fleischer, Eur. Phys. J. **C6** (1999) 647.

[156] R. Fleischer, Phys. Rev. **D60** (1999) 073008.

[157] A. Belkov, S. Shulga, Part. Nucl. Lett. **117** (2003) 11; Comput. Phys. Commun. **156** (2004) 221.

[158] R. Fleischer, Phys. Lett. **B562** (2003) 234.

[159] R. Fleischer, Nucl. Phys. **B659** (2003) 321.

[160] M. Gronau and J.L. Rosner, Phys. Lett. **B482** (2000) 71.

[161] P.Z. Skands, JHEP **0101** (2001) 008.

[162] G. Balbi *et al.*, CERN-LHCb/2003-123 and 124;
R. Antunes Nobrega *et al.* [LHCb Collaboration], *Reoptimized LHCb Detector, Design and Performance*, Technical Design Report 9, CERN/LHCC 2003-030.

[163] A. Khodjamirian, T. Mannel and M. Melcher, Phys. Rev. **D68** (2003) 114007.

[164] M. Beneke, eConf **C0304052** (2003) FO001 [hep-ph/0308040].

[165] R. Fleischer, Eur. Phys. J. **C16** (2000) 87.

[166] A.J. Buras, F. Parodi and A. Stocchi, JHEP **0301** (2003) 029.

[167] K. Abe *et al.* [Belle Collaboration], BELLE-PREPRINT-2004-1 [hep-ex/0401029].

[168] R. Aleksan, I. Dunietz and B. Kayser, Z. Phys. **C54** (1992) 653.

[169] I. Dunietz and R.G. Sachs, Phys. Rev. **D37** (1988) 3186 [E: **D39** (1989) 3515];
I. Dunietz, Phys. Lett. **B427** (1998) 179;
D.A. Suprun, C.W. Chiang and J.L. Rosner, Phys. Rev. **D65** (2002) 054025.

[170] R. Fleischer, Nucl. Phys. **B671** (2003) 459.

[171] B. Aubert *et al.* [BaBar Collaboration], BABAR-PUB-03-033 [hep-ex/0310037].

[172] K. Abe *et al.* [Belle Collaboration], BELLE-CONF-0341 [hep-ex/0308048].

[173] A.J. Buras and M. Misiak, Acta Phys. Polon. **B33** (2002) 2597.

[174] G. Buchalla, A.J. Buras and M.K. Harlander, Nucl. Phys. **B349** (1991) 1.

[175] A.J. Buras and M.K. Harlander, Adv. Ser. Direct. High Energy Phys. **10** (1992) 58.

[176] A. Ali, CERN-TH/2002-284 [hep-ph/0210183];
T. Hurth, Rev. Mod. Phys. **75** (2003) 1159;
K. Bieri and C. Greub, hep-ph/0310214.

[177] G. Buchalla and A.J. Buras, Nucl. Phys. **B548** (1999) 309.

[178] G. Buchalla and A.J. Buras, Nucl. Phys. **B400** (1993) 225.

[179] M. Misiak and J. Urban, Phys. Lett. **B451** (1999) 161.

[180] D. Acosta *et al.* [CDF Collaboration], FERMILAB-PUB-036-E [hep-ex/0403032];
M.C. Chang *et al.* [Belle Collaboration], Phys. Rev. **D68** (2003) 111101.

[181] A.J. Buras, Phys. Lett. **B566** (2003) 115.

[182] A.J. Buras, TUM-HEP-544-04 [hep-ph/0402191].

[183] G. Buchalla and A.J. Buras, Nucl. Phys. **B398** (1993) 285.

[184] A.J. Buras and L. Silvestrini, Nucl. Phys. **B546** (1999) 299.

[185] A.J. Buras, G. Colangelo, G. Isidori, A. Romanino, L. Silvestrini, Nucl. Phys. **B566** (2000) 3;

A.J. Buras, A. Romanino and L. Silvestrini, Nucl. Phys. **B520** (1998) 3.

[186] G. Buchalla, G. Hiller and G. Isidori, Phys. Rev. **D63** (2001) 014015;
D. Atwood and G. Hiller, hep-ph/0307251.

[187] V. Barger, C.W. Chiang, P. Langacker and H.S. Lee, Phys. Lett. **B580** (2004) 186.

[188] J. Kaneko *et al.* [Belle Collaboration], Phys. Rev. Lett. **90** (2003) 021801;
B. Aubert *et al.* [BaBar Collaboration], BABAR-CONF-03-19 [hep-ex/0308016].

[189] A. Alavi-Harati *et al.* [KTeV Collaboration], FERMILAB-PUB-03-446 [hep-ex/0309072].

[190] Y. Grossman and Y. Nir, Phys. Lett. **B398** (1997) 163.

[191] G. Buchalla, G. D'Ambrosio and G. Isidori, Nucl. Phys. **B672** (2003) 387.

[192] G. Isidori, C. Smith and R. Unterdorfer, hep-ph/0404127.

[193] S.R. Choudhury, N. Gaur and A.S. Cornell, hep-ph/0402273.

[194] P.H. Chankowski, J.R. Ellis, S. Pokorski, M. Raidal and K. Turzynski, CERN-PH-TH/2004-030 [hep-ph/0403180].

BEYOND THE STANDARD MODEL: EXTRA DIMENSIONS AND SUPERSYMMETRY

G. Gabadadze
Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY, 10003

Abstract
I give an elementary introduction to two possible extensions of the Standard Model of particle physics. The first one assumes the existence of extra dimensions and the second one of a supersymmetric world. I outline the basic principles of the theories with extra dimensions and/or supersymmetry and discuss certain phenomenological consequences of these models.

1 INTRODUCTION

The Standard Model (SM) of strong and electroweak interactions is an extremely successful model that *parametrizes* all the existing particle physics data with an extraordinary accuracy (for a recent, brief status review see, for example, Ref. [1]).

However, the SM does not give an *explanation* of many interesting particle properties and interactions that it describes with such success. To name just one, the origin of fermion mass patterns, the reason for the existence of the three generations, charge quantization, and certain properties that make the SM consistent with the observations, cannot be understood within the SM itself.

Moreover, some cosmological and astrophysical issues, such as the baryon asymmetry of the Universe, the problem of Dark Matter, and the origin of ultra-high-energy cosmic rays, are hard to understand without invoking certain new physics beyond the SM. These issues were covered in detail by I. Tkachev in this school [2].

Based on one or all of the above argumentations, physicists are searching for theories beyond the Standard Model. A new, successful model has to include in it all the ingredients of the SM and has to go beyond the SM in explaining the mysteries that cannot be explained by the SM.

During the last 30 years or so, this search was mainly driven by the following two theoretical concepts:

- Unification of strong and electroweak interactions
- The hierarchy problem

The unification is an extremely powerful concept that explores the possibility that the strong, electromagnetic, and weak interactions have a common origin [3]. A typical unification takes place at energies of the order of 10^{16} GeV.

The hierarchy problem is the theorist's dissatisfaction with the fact that the Higgs mass in the SM is very sensitive to high-energy physics because of the quantum loops. A tremendous fine-tuning is needed in order to keep the Higgs mass light enough for it to be relevant for electroweak symmetry breaking.

Attempts to solve the hierarchy (or Higgs mass) problem gave rise to at least three major directions beyond the SM that have been explored in some detail. Each of these directions contains a number of interesting and elaborate models that by themselves introduce new concepts and give rise to new consequences, so putting them in the three directions is convenient but not very fair. These models typically predict new physics in a few-TeV region that could be tested at the LHC!

In an arbitrary order these directions are

– Supersymmetry
– Large extra dimensions
– Composite/Nambu–Goldstone Higgs models

These lectures will give an introduction to the first two approaches—large extra dimensions and supersymmetry. Because of lack of time, the third approach was not covered at this School. This is unfortunate since there were interesting recent developments in the above direction. The reader is referred to the literature [4].

2 THE HIERARCHY PROBLEM

In this section we review briefly what is called the hierarchy problem. Any parameter in the SM Lagrangian can potentially get renormalized due to quantum loop corrections. The particle masses are among these parameters. The masses of the SM fermions and gauge bosons are protected from large quantum-loop renormalizations by symmetries of the theory. However, there is no such symmetry for a scalar Higgs particle in the SM.

Let us parametrize the renormalized Higgs mass as follows:

$$\left(M_{\rm H}^{\rm Ren}\right)^2 \,=\, \left(M_{\rm H}^{0}\right)^2 \,+\, \delta M_{\rm H}^2 \,\leq\, \left({\rm TeV}\right)^2 \,. \tag{1}$$

Here $(M_{\rm H}^0)^2$ denotes the bare Higgs mass that enters the bare SM Lagrangian. In order for the Higgs particle to be relevant for Electroweak (EW) symmetry-breaking, the renormalized Higgs mass has to be $\lesssim$ TeV.

However, $\delta M_{\rm H}^2$ is power-sensitive to ultraviolet (UV) physics. In particular, if $\Lambda_{\rm UV}$ is a UV cutoff of the theory at hand, then we get

$$\delta M_{\rm H}^2 \,\sim\, \int_0^{\Lambda_{\rm UV}} \frac{d^4k}{k^2} \,\sim\, \Lambda_{\rm UV}^2 \,. \tag{2}$$

Furthermore, if one assumes that there is no new physics all the way up to the Planck scale, then $\Lambda_{\rm UV} \sim M_{\rm Pl}$,and

$$\left(M_{\rm H}^{\rm Ren}\right)^2 \gg \left({\rm TeV}\right)^2 \,. \tag{3}$$

Such a high Higgs mass is not acceptable. The above arguments indicate that a possible new physics can be entering at much lower energies, $\Lambda_{\rm UV} \ll M_{\rm Pl}$, rendering the Higgs light.

There are the following possibilities:

(i) Higgs is a composite state with the compositness scale $\sim$TeV. Then, $\Lambda_{\rm UV}$ $\sim$TeV, and there is no problem.

(ii) Supersymmetry (SUSY). Typically, in this case $\Lambda_{\rm UV} \sim 10^{16}$ GeV, however, SUSY is broken at a low scale, $\sim$ TeV, and the SUSY particles are entering the game at around a TeV.

(iii) Large extra dimensions. One declares that $\Lambda_{\rm UV} \sim M_{\rm Quantum\,Gravity} \sim$ TeV. This nullifies the hierarchy problem.

(iv) Yet another possibility is to assume that $\Lambda_{\rm UV} \sim M_{\rm Pl}$ and to accept a fine-tuning of 1 part in 10^{15} that is needed to arrange for the cancellation between the bare Higgs mass and $\delta M_{\rm H}^2$.

In what follows we shall explore the consequences of (iii) and (ii).

3 INTRODUCTION TO EXTRA DIMENSIONS

Extra dimensions had been studied long before the SM and particle physics emerged. They were introduced with the aim of unifying the gravitational and electromagnetic interactions. In this section we shall

more or less follow the historical course of the development of theories with extra dimensions (in this section we follow Ref. [5]).

The magnitude of gravitational force F between two macroscopic objects separated at a distance r obeys the inverse-square law, $F \sim r^{-2}$. This would not be so if the world had $N \geq 1$ extra spatial dimensions that are similar to our three—in that case we would instead measure $F \sim r^{-(2+N)}$. Similar arguments hold for the microworld of elementary particles. For instance, we know from accelerator experiments that electromagnetic interactions of charged particles obey the inverse-square law.

However, experimental capabilities are limited and so is our knowledge of the validity of these laws of nature. For instance, it has not been established how gravity behaves at distances shorter than 10^{-4} cm, or at distances larger than 10^{28} cm. All we know is that for $10^{-4}\text{cm} \lesssim r \lesssim 10^{28}$ cm the inverse-square law provides a good description of nonrelativistic gravitational interactions, but laws of nature might be different outside of that interval. Likewise, we are certain that electromagnetic interactions obey the inverse-square law all the way down to distances of order 10^{-16} cm, but they might change somewhere below that scale.

At present, it is not clear how exactly these laws of nature might change. There is a possibility that they will change according to the laws of higher-dimensional space if extra dimensions exist. However, it is fair to wonder why one should think in the first place that the world might have extra dimensions. I shall give below major theoretical arguments that motivated an enormous amount of research in the field of extra dimensions.

The first scientific exploration of the idea of extra dimensions was by Kaluza [6] and Klein [7]. They noticed that gravitational and electromagnetic interactions, since they are so alike, could be descendants of a common origin. However, amazingly enough, the unified theory of gravity and electromagnetism could be formulated only in space with extra dimensions. Subsequently, non-Abelian gauge fields, similar to those describing weak and strong interactions, were also unified with Einstein's gravity in models with extra dimensions. Therefore, the first reason why extra dimensions were studied was:

– Unification of gravity and gauge interactions of elementary particles.

So far we have been discussing classical gravitation. However, quantization of gravity is a very nontrivial task. A candidate theory of quantum gravity, string theory (M-theory), can be formulated consistently in space with extra six or seven dimensions; hence, the second reason to study extra dimensions:

– Quantization of gravitational interactions.

All the extra dimensions considered above were very small, of the Planckian size and therefore undetectable. A new wave of activity in the field of extra dimensions came with the framework of Arkani-Hamed, Dimopoulos and Dvali (ADD) [8] who observed that the Higgs mass hierarchy problem can be addressed in models with *large extra dimensions*. Because the extra dimensions are large in the ADD framework, their effects can be measurable in future accelerator, astrophysical, and table-top experiments. Moreover, these models can be embedded in a string theory framework [9]. Subsequently Randall and Sundrum proposed a model with warped extra dimension [10] that also provides an attractive set-up for addressing the Higgs mass hierarchy problem and for studying physical consequences of extra dimensions. Thus, the third reason is:

– The Higgs mass hierarchy problem.

Another type of hierarchy problem is the problem of the cosmological constant. The latter is very hard to address unless one of the conventional notions such as locality, unitarity, causality, or four-dimensionality of space–time is given up. In that regard, theories with *infinite volume* extra dimensions [11]—the only theories that are not four-dimensional at very low energies—were proposed as a candidate for solving the cosmological constant problem [12, 13]. Hence the fourth reason is:

– The cosmological constant problem.

In what follows I shall discuss some of the developments in extra dimensional theories listed above.

4 INTRODUCTION TO KALUZA–KLEIN THEORIES

Extra spatial dimensions are not similar to our three dimensions in the Kaluza–Klein (KK) approach. Instead, the extra dimensions form a *compact* space with a certain compactification scale L. For instance, one extra dimension can be a circle of radius L, or simply an interval of size L. For more than one extra dimension this space could be a higher dimensional sphere, torus, or some other manifold. In general, D-dimensional space–time in the KK approach has a geometry of a direct product $M^4 \times X^{D-4}$ where M^4 denotes four-dimensional Minkowski space–time, and X^{D-4} denotes a compact manifold of extra dimensions—called an *internal manifold*[1].

In the KK approach there is a certain dynamics in D-dimensional space–time that gives rise to preferential compactification of the extra $(D-4)$-dimensions leaving four Minkowskian dimensions intact. The geometry $M^4 \times X^{D-4}$ should be a solution of D-dimensional Einstein equations.

Let us now discuss the physical implications of the compact extra dimensions. Based on common sense, it is clear that at distance scales much larger than L, the extra dimensions should not be noticeable. They only become 'visible' when one probes very short distances of order L.

To discuss these properties in detail we start with the simplest example of a real scalar field in $(4+1)$-dimensional space–time. In the the paper we use the mostly positive metric $[-+++{+}..]$. The Lagrangian density takes the form

$$\mathcal{L} \,=\, -\frac{1}{2}\,\partial_A\,\Phi\partial^A\,\Phi\,, \qquad A=0,1,2,3,5\,. \tag{4}$$

Here the field $\Phi(t,\vec{x},y) \,\equiv\, \Phi(x_\mu,y)\,,\;\; \mu=0,1,2,3$, depends on four-dimensional coordinates x_μ as well as on an extra coordinate y. The extra dimension is assumed to be compactified on a circle S^1 of radius L. Therefore, the five-dimensional space–time has a geometry of $M^4 \times S^1$. In this space the scalar field should be periodic with respect to $y \to y + 2\pi L$:

$$\Phi(x,y) \,=\, \Phi(x,y+2\pi L)\,. \tag{5}$$

Let us now expand this field in the harmonics on a circle

$$\Phi(x,y) \,=\, \sum_{n=-\infty}^{+\infty}\,\phi_n(x)\,e^{iny/L}\,. \tag{6}$$

[Note that $\phi_n^*(x) = \phi_{-n}(x)$.] Substituting this expansion into Eq. (4) the Lagrangian density (4) can be rewritten as follows

$$\mathcal{L} \,=\, -\frac{1}{2}\sum_{n,m=-\infty}^{+\infty}\left(\partial_\mu\phi_n\partial^\mu\phi_m \,-\, \frac{nm}{L^2}\,\phi_n\phi_m\right)e^{i(n+m)y/L}\,, \tag{7}$$

while the action takes the form

$$S \,=\, \int d^4x\int_0^{2\pi L}dy\mathcal{L} \,=\, -\frac{2\pi L}{2}\int d^4x\sum_{n=-\infty}^{+\infty}\left(\partial_\mu\phi_n\partial^\mu\phi_n^* \,+\, \frac{n^2}{L^2}\phi_n\phi_n^*\right)\,. \tag{8}$$

On the right-hand side of the above equation we performed integration w.r.t. y. The resulting expression is an action for an infinite number of four-dimensional fields $\phi_n(x)$. To study properties of these fields it is convenient to introduce the notation

$$\varphi_n \equiv \sqrt{2\pi L}\phi_n\,. \tag{9}$$

[1]The X^{D-4} does not have to be a manifold in a strict mathematical definition of this notion (see examples below), however, we shall use this name most of the time for the sake of simplicity.

The latter allows us to rewrite the action in the following form

$$S = \int d^4x \left[-\frac{1}{2}\,\partial_\mu\varphi_0\partial^\mu\varphi_0 \right] - \int d^4x \sum_{k=1}^{+\infty} \left(\partial_\mu\phi_k\partial^\mu\phi_k^* + \frac{k^2}{L^2}\phi_k\phi_k^* \right) . \tag{10}$$

Therefore, the spectrum of a compactified theory consists of:

- A single, real, massless scalar field, called a *zero-mode*, φ_0;
- An infinite number of massive, complex, scalar fields with masses inversely proportional to the compactification radius, $m_k^2 = k^2/L^2$.

All the states mentioned above are called the Kaluza–Klein modes. At low energies, i.e., when $E \ll 1/L$, only the zero mode is important; while at higher energies $E \gtrsim 1/L$, all the KK modes become essential.

As a next step we consider a $(4+1)$-dimensional example of Abelian gauge fields. An additional ingredient, compared to the scalar case, is the local gauge invariance, the consequences of which we shall emphasize below.

Let us start with the Lagrangian density

$$\mathcal{L} = -\frac{1}{4g_5^2}\,F_{\rm AB}F^{\rm AB}\,, \tag{11}$$

where the dimensionalities are set as follows: $[A_B] = [mass]$, $[g_5^{-2}] = [mass]$. As in the previous example, we assume compactification on a circle S^1 of radius L and periodic boundary conditions on the fields. We decompose $F_{\rm AB}^2 = F_{\mu\nu}^2 + 2(\partial_\mu A_5 - \partial_5 A_\mu)^2$, and expand the fields A_μ and A_5 in the harmonics on a circle

$$A_\mu(x,y) = \sum_{n=-\infty}^{+\infty} A_\mu^{(n)}(x)\, e^{iny/L}\,, \quad A_5(x,y) = \sum_{n=-\infty}^{+\infty} A_5^{(n)}(x)\, e^{iny/L}\,. \tag{12}$$

As in the scalar example, we integrate w.r.t. y to calculate the effective 4D action

$$S = \int d^4x \int_0^{2\pi L} dy\mathcal{L} \equiv \int d^4x\, \mathcal{L}_4\,. \tag{13}$$

Using gauge transformation, the expression for $\mathcal{L}_4$ can be cast in the following form

$$\mathcal{L}_4 = -\frac{1}{4g_4^2} \left\{ F_{\mu\nu}^{(0)}F^{(0)\mu\nu} + 2\sum_{k=1}^{+\infty} \left[F_{\mu\nu}^{(k)}F^{*(k)\mu\nu} + \frac{2k^2}{L^2}\,A_\mu^{(k)}A^{*(k)\mu} \right] + 2(\partial_\mu A_5^{(0)})^2 \right\} . \tag{14}$$

Therefore, we conclude that the spectrum of the compactified model consists of the following states:

- A zero-mode—a massless gauge field $A_\mu^{(0)}$ with the gauge coupling $g_4^2 = g_5^2/(2\pi L)$;
- Massive KK gauge bosons with the mass $m_k^2 = k^2/L^2$;
- Massless scalar field $A_5^{(0)}$.

A few words on local gauge invariance are in order here. The five-dimensional model is invariant under five-dimensional local gauge transformations $A_B(x,y) \to A_B(x,y) + \partial_B\alpha(x,y)$. After compactification the five-dimensional gauge transformations reduce to an infinite number of *four-dimensional* gauge transformations—one for each KK level $A_\mu^{(n)}(x) \to A_\mu^{(n)}(x) + \partial_\mu\alpha^{(n)}(x)$. However, only the zero-mode is a massless gauge field, all the higher KK modes are massive. This can be interpreted as a consequence of the Higgs mechanism taking place on each massive KK level where a massless gauge field 'eats' one massless scalar $A_5^{(n)}$ and becomes a massive gauge field with three physical degrees of

freedom. On the massless level there is a 4D massless gauge field with two physical degrees of freedom plus one real massless scalar $A_5^{(0)}$.

Finally we come to the main subject of this section and consider a $(4+1)$-dimensional example of gravity. It demonstrates how 4D Einstein gravity can be unified with electromagnetism in a 5d theory — the original proposal of Kaluza and Klein.

The 5d action takes the form

$$S = \frac{M_*^3}{2} \int d^4x dy \sqrt{G} R_5 \,. \tag{15}$$

As in the previous examples, the space is $M^{(4)} \times S^1$ and we expand fields in the harmonics on a circle of radius L

$$G_{\rm AB}(x,y) \;=\; \sum_{n=-\infty}^{+\infty} G_{\rm AB}^{(n)}(x)\, e^{iny/L} \,. \tag{16}$$

In what follows we shall concentrate on the zero mode $G_{\rm AB}^{(0)}$ neglecting all the massive modes.

Let us introduce the notations

$$\begin{aligned} G_{\mu\nu}^{(0)} = e^{\phi/\sqrt{3}}(g_{\mu\nu}(x) + e^{-\sqrt{3}\phi} A_\mu A_\nu) \,, \\ G_{\mu 5}^{(0)} = G_{5\mu}^{(0)} = e^{-2\phi/\sqrt{3}} A_\mu \,, \\ G_{55}^{(0)} = e^{-2\phi/\sqrt{3}} \,. \end{aligned} \tag{17}$$

Using these expressions we find the 4D action for the zero-mode fields

$$S_{\rm zm} \;=\; M_*^3 \pi L \int d^4x \sqrt{g} \left(R_4(g) - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{4} e^{-\sqrt{3}\phi} F_{\mu\nu}^2 \right) . \tag{18}$$

Recalling that the conventional 4D action for gravity has a form

$$\frac{M_{\rm Pl}^2}{2} \int d^4x \sqrt{g} R_4(g) \,, \tag{19}$$

we find that $M_{\rm Pl}^2 = M_*^3 2\pi L$. As a result, the Newton constant $G_{\rm N} = (8\pi M_{\rm Pl}^2)^{-1}$ can be related to the higher dimensional scale and the compactification radius

$$G_{\rm N} = \frac{1}{16\pi^2 M_*^3 L} \,. \tag{20}$$

The main result of the above discussion is that four-dimensional gauge and gravitational fields have a common origin in five-dimensional gravitational field.

Let us count physical degrees of freedom. A four-dimensional massless graviton has two physical degrees of freedom (pdf's); A four-dimensional massless gauge boson has also two pdf's, and a real scalar has one pdf. The total is five pdf's, in agreement with five pdf's of a massless five-dimensional graviton[2].

Let us now turn to the massive KK levels. The analysis is similar to that of gauge fields but more cumbersome. Nevertheless, the main results can be summarized as follows. There is a massive graviton with the mass $m_k^2 = k^2/L^2$ at each k'th level. These gravitons acquire masses via the Higgs

[2]In general, the total number of independent components of a rank 2 symmetric tensor in D-dimensions is $D(D+1)/2$, however, only $D(D-3)/2$ of those correspond to physical degrees of freedom of a D-dimensional massless graviton; the remaining extra components are the redundancy of manifestly gauge- and Lorentz-invariant descriptions of the theory.

mechanism—one massless graviton (two pdf's) 'eats' one massless gauge boson (two pdf's) and one real scalar (one pdf)—this makes one massive 4D graviton that has five pdf's. The massive gravitational KK modes are charged under the massless gauge field. The charges are determined as $q_k \sim k/LM_{\rm Pl} \sim m_n/M_{\rm Pl}$. At the linearized level, gauge transformations do not mix with other different KK levels, however, this mixing shows up once the nonlinear interactions of gravitational theory are taken into account [14].

5 INTRODUCTION TO BRANEWORLDS

The idea that our $(3+1)$-dimensional world could be realized as a 3d surface in higher dimensional space was actively discussed in the context of general relativity in the 1960s and 1970s.

A first, particle physics application of this idea was put forward by Rubakov and Shaposhnikov [15] and independently by Akama [16].

In this section, following Ref. [15], we consider a toy example of the braneworld where the main mechanism of localization can be worked out explicitly.

We start with a scalar field in five dimensions with the following Lagrangian density

$$\mathcal{L} \,=\, -\frac{1}{2}\,\partial_A\,\Phi\partial^A\,\Phi \,-\, \frac{\lambda}{2}\,\left(\Phi^2-\eta^3\right)^2\,. \tag{21}$$

The Lagrangian is invariant under the $\mathbf{Z}_2$ transformations $\Phi \to -\Phi$, however, the vacua of the theory are not — under the $\mathbf{Z}_2$ the two vacua $\Phi = \pm\eta^{3/2}$ interchange. Therefore, the $\mathbf{Z}_2$ is spontaneously broken. As a result, there should exist domain walls. We find the following domain wall (kink) solution to the classical equation of motion

$$\Phi_{\rm cl}(y) \,=\, \eta^{3/2}{\rm tanh}\left(\sqrt{\lambda}\eta^{3/2}\,y\right) \,\equiv\, \eta^{3/2}{\rm tanh}\left(m_0\,y\right)\,. \tag{22}$$

Transverse to the domain wall, space is one-dimensional, hence, the domain wall is a codimension-one object. Its worldvolume has three spatial coordinates, therefore, it is also called a 3-brane.

Let us discuss certain properties of the solution. The tension of the wall is its surface energy density $T = \int dy H(\Phi_{\rm cl}) = \int dy T_{00}(\Phi_{\rm cl})$, where H denotes the Hamiltonian and T_{00} denotes the 00 component of the stress tensor. The tension is determined as follows

$$T \,\sim\, \frac{m_0^3}{\lambda} \,\sim\, \sqrt{\lambda}\eta^{3/2}\eta^3\,. \tag{23}$$

Below we would like to understand what are the excitations that live on the brane worldvolume. According to the braneworld idea [15], [16], in a realistic construction, those excitations should be identified with the Standard Model particles. For this purpose we perform the following decomposition

$$\Phi(x,y) \,=\, \Phi_{\rm cl}(y) \,+\, \delta\Phi(x,y)\,. \tag{24}$$

Then we find that the 5d equations have a solution

$$\delta\Phi(x,y) \,=\, \left(\frac{d\Phi_{\rm cl}}{dy}\right)\,\rho(x)\,, \tag{25}$$

where the four-dimensional field ρ satisfies the equation

$$\partial_\mu^2\rho = 0\,. \tag{26}$$

Therefore, ρ is nothing but a massless four-dimensional mode. The wavefunction of this mode is proportional to $d\Phi_{\rm cl}/dy$ and vanishes outside of the brane. Therefore, this mode is localized on a brane. This

excitation is just a Nambu–Goldstone boson of spontaneously broken translation invariance along the y direction.

Let us now introduce fermions. For this we add to the Lagrangian the following two terms

$$\Delta\mathcal{L} \;=\; i\bar{\Psi}\Gamma^M\partial_M\Psi - h\Phi\bar{\Psi}\Psi\,, \tag{27}$$

where Ψ denotes a 5-dimensional Dirac fermion. The equation of motion for the fermion in the background of the domain wall reads as follows:

$$i\Gamma^M\partial_M\Psi - h\Phi_{\rm cl}\Psi = 0\,. \tag{28}$$

This equation has a normalizeable solution of the following form

$$\Psi_{\rm zm}(x,y) \;=\; e^{-\int_0^y h\Phi_{\rm cl}(z)dz}\chi_L(x)\,, \tag{29}$$

where χ_L denotes a four-dimensional massless chiral mode

$$i\Gamma^\mu\partial_\mu\chi_L = 0, \quad \chi_L = (1-\gamma_5)\chi/2\,. \tag{30}$$

From this expression we see that the wavefunction of this mode vanishes outside of the brane. Therefore, one obtains a four-dimensional chiral mode that is localized on the worldvolume[3].

Summarizing, in a simple construction described above, scalars and fermions can be localized on a brane. However, for realistic model building one should in addition perform two major steps:

(i) Localize gauge fields on a brane;

(ii) Obtain four-dimensional gravity on the brane.

A mechanism for gauge field localization within the field theory context was proposed by Dvali and Shifman [17]. It is based on the observation that a gauge field can be in the confining phase, the bulk, while being in the broken phase on a brane; then confining potential prevents the low-energy brane gauge fields from propagating into the bulk. This mechanism is discussed in detail in Ref. [17].

Localization of gauge fields is a rather natural property of D-branes in closed string theories [18]—the gauge fields emerge on a brane as fluctuations of open strings that are attached to the brane and do not exist in the bulk.

As to issue (ii), we discuss below three distinct mechanisms by means of which the laws of 4D gravity can be obtained on a brane.

6 BRANEWORLDS WITH COMPACT EXTRA DIMENSIONS

One way to obtain 4D gravity on a brane is to combine the braneworld idea with the idea of KK compactification. This, as was proposed by Arkani-Hamed, Dimopoulos and Dvali (ADD) [8], opens up new possibilities to solve the Higgs mass hierarchy problem and gives rise to new predictions that can be tested in accelerator, astrophysical and table-top experiments. Moreover, the framework can be embedded in string theory [9].

The main ingredients of the simplest ADD scenario are:

- Standard Model particles are localized on a 3-brane, while gravity spreads to all $4+N$ dimensions.
- The fundamental scale of gravity M_*, and the ultraviolet (UV) scale of the Standard Model, are around a few TeV or so. This can eliminate the Higgs mass hierarchy problem.
- N extra dimensions are compactified.

[3]There also exists a solution with an opposite chirality that is not localized on a brane.

The action for the simplest ADD model takes the form:

$$S_{\rm ADD} = \frac{M_*^{2+N}}{2} \int d^4x \int_0^{2\pi L} d^N y \sqrt{G} R_{(4+N)} + \int d^4x \sqrt{g} (T + \mathcal{L}_{\rm SM}(\Psi, M_{\rm SM})) \tag{31}$$

where $M_* \sim (1-10)$ TeV, $g(x) = G(x, y=0)$, $T + \langle \mathcal{L}_{\rm SM} \rangle = 0$, the latter condition is a usual fine-tuning of the cosmological constant.

Technical simplifications which are adopted above but that can be easily lifted are as follow:

(1) The brane width is taken to be zero (generically, the natural scale for the brane width could be M_*^{-1}).

(2) Brane fluctuations are neglected (these are Nambu–Goldstone bosons which couple to matter derivatively).

(3) All extra dimensions have equal size L (in general, different extra dimensions could have different sizes).

(4) Only gravity can propagate in the bulk (in general, other fields could also live in the bulk; in fact there are attractive scenarios with right-handed neutrinos living in the bulk [19]).

Let us first study the properties of 4D gravity in the ADD scenario. The low, effective 4D action for a zero mode takes the form

$$\frac{M_*^{2+N}}{2} \int d^4x \int_0^{2\pi L} d^N y \sqrt{G} R_{(4+N)} \to \frac{M_*^{2+N} (2\pi L)^N}{2} \int d^4x \sqrt{g_{\rm zm}} R_{\rm zm} \,, \tag{32}$$

hence, we should define the 4D Planck mass

$$M_{\rm Pl}^2 = M_*^{2+N} (2\pi L)^N \,. \tag{33}$$

Postulating that the quantum gravity scale is at $M_* \sim$ TeV we find what should be the size of extra dimensions

$$L \sim 10^{-17+30/N} \,{\rm cm} \,. \tag{34}$$

For one extra dimension, $N=1$, one gets $L \sim 10^{13}$ cm, this is excluded within the ADD framework since gravity below 10^{13} would have been higher dimensional. For $N=2$ we get $L \sim 10^{-2}$ cm; this particular case is very interesting since it predicts modification of the 4D laws of gravity at submillimetre distances—the subject of active experimental studies. For larger N the value of L should decrease; but even for $N=6$, L is very large compared to $1/M_{\rm Pl}$.

Two static sources on the brane interact with the following nonrelativistic gravitational potential

$$V(r) = -G_N m_1 m_2 \sum_{n=-\infty}^{+\infty} |\Psi_n(y=0)|^2 \frac{e^{-m_n r}}{r} \,, \tag{35}$$

where $\Psi_n(y=0)$ denotes the wavefunction of the n-th KK mode at a position of a brane and $m_n = |n|/L$. If $r \gg L$, from the above expression we find

$$V(r) = -\frac{G_N m_1 m_2}{r} \,. \tag{36}$$

This recovers the conventional 4D law of Newtonian dynamics. In the opposite limit, i.e., when $r \ll L$ one gets

$$V(r) = -\frac{m_1 m_2}{M_*^{2+N} r^{1+N}} \,. \tag{37}$$

That is the law of $(4+N)$-dimensional gravitational interactions. Therefore, the laws of gravity are modified at distances of order L.

Selected topics of the ADD phenomenology:

– *Gauge coupling unification.* In a conventional 4D theory the renormalization group running of the gauge coupling constants is logarithmic. This changes in higher dimensions where the power-law running takes place [20]. As was shown by Dienes, Dudas and Gherghetta [21], the power-law running is what gives rise to *an accelerated unification* of the strong, weak and electromagnetic couplings at a scale around M_* in braneworlds with compact extra dimensions.

– *Missing energy signals in accelerator experiments.* The SM particles are localized on a brane only up to some energy scale that is comparable to M_*. At about that scale the SM particles could in principle escape into the bulk. This would provide missing energy signals in accelerator experiments. Another missing energy signal can be due to emission of KK gravitons into the bulk, see detailed discussions in Refs. [8], [22].

– *Energy loss by stars via emission of light KK gravitons.* In the 6d ADD model the KK gravitons are light, $m_{KK} \sim L^{-1} \sim 10^{-4}$eV. Therefore, these gravitons can be emitted in the interior of astrophysical objects the temperature of which exceeds 10^{-4} eV. As a result, these objects, such as stars, can cool down due to the process of emission of the KK gravitons into the bulk. Each KK graviton emission is $M_{\rm Pl}$ suppressed. However, because of the the high-multiplicity of the KK graviton, the net result for the emission rate is suppressed by $1/M_*^2$. Unless this rate is small enough, a star would cool down faster than it should by emitting these KK gravitons. This puts a lower bound on M_* in a 6d theory to be 50 TeV or so [8, 23].

– *Cosmological implications.* There exist new scenarios of inflation and baryogenesis within the braneworld context. These scenarios manifestly use properties of branes. For instance, inflation on 'our brane' can be obtained if another brane falls on top of 'our brane' in the early period of development of the brane-universe [24]. The potential that is created by another brane in 'our world' can be viewed as the conventional inflationary potential. Baryon asymmetry of a desired magnitude can also be produced during the collision of these two branes [25]. For more recent developments see Refs. [26]– [29].

6.1 Phenomenology of large extra dimensions

In this subsection we shall discuss one representative example of a particle reaction with missing energy that escapes into extra dimensions. We consider the $e^+\,e^-$ annihilation into a photon and KK graviton $e^+\,e^- \to \gamma\, G_m$. The differential cross section takes the form [22]:

$$\frac{d\sigma_m}{dt} = \frac{\alpha}{16}\frac{1}{s\,M_{\rm Pl}^2}F_1(t/s, m^2/s)\,, \tag{38}$$

where

$$F_1(a,b) = \frac{1}{a(b-a-1)}\left(-4a(1+a)(1+2a+2a^2) + b(1+6a+18a^2+16a^3)\right) \\ \frac{1}{a(b-a-1)}\left(-6b^2a(1+2a) + b^3(1+4a)\right). \tag{39}$$

Notice the $1/M_{\rm Pl}^2$ suppression of the differential cross section. However, this suppression will go away when we sum up the large multiplicity of the KK states G_m. Indeed, the cutoff of the theory is M_*, and the KKs have masses $\sim 1/L$. The maximal number of KKs is then $n_{\rm KK}^{\rm max} \sim (M_*L)^N$. Because $M_*^{2+N}L^N \sim M_{\rm Pl}^2$ we find $n_{\rm KK}^{\rm max} \sim M_{\rm Pl}^2/M_*^2$. Since $M_* >> 1/L$ we obtain

$$\sum_{\rm KK} \to (2\pi L)^N \int \frac{d^N q_\perp}{(2\pi)^N} = (2\pi L)^N \Omega_N \int_0^{M_*} \frac{q_\perp^{N-1} dq_\perp}{(2\pi)^N}\,, \tag{40}$$

where

$$\Omega_N = \frac{2\pi^{N/2}}{\Gamma(N/2)}\,. \tag{41}$$

Finally, the sum can be replaced by the integral

$$\sum_{\rm KK} \to \frac{M_{\rm Pl}^2}{M_*^{2+N}}\Omega_N \int_0^{M_*} \frac{m^{N-1}dm}{(2\pi)^N}\,. \tag{42}$$

Using this fact we obtain

$$\frac{d^2\sigma_m}{dt\,dm} \sim \frac{\alpha}{s\,M_*^{2+N}} m^{N-1} F_1(t/s, m^2/s)\,. \tag{43}$$

It is convenient to introduce the notations: $z_1 \equiv t/s$ and $z_2 \equiv m/\sqrt{s}$. With these new variables we find:

$$\frac{d^2\sigma_m}{dz_1\,dz_2} \sim \frac{\alpha s^{N/2}}{M_*^{2+N}} z_2^{N-1} F_1(z_1, z_2^2)\,. \tag{44}$$

The $1/M_{\rm Pl}^2$ suppression is gone! As was pointed out before, this is due to the large multiplicity of the KK states.

Similar considerations apply to the following processes

$$\mathrm{p\bar{p}} \to \text{jet} + \text{missing energy}\,; \quad \mathrm{pp} \to \text{jet} + \text{missing energy}\,. \tag{45}$$

The above reactions can arise from the following parton subprocesses

$$\mathrm{q\bar{q}} \to \mathrm{Gg}, \quad \mathrm{g\bar{q}} \to \mathrm{G\bar{q}}, \quad \mathrm{gg} \to \mathrm{gG}, \quad \mathrm{qg} \to \mathrm{qG}\,. \tag{46}$$

As can be shown, the differential cross section takes the form

$$\frac{d\sigma_m}{dt}|\mathrm{q\bar{q}} \to \mathrm{Gg} = \frac{\alpha}{36}\frac{1}{s\,M_{\rm Pl}^2} F_1(t/s, m^2/s)\,. \tag{47}$$

As before, the $1/M_{\rm Pl}^2$ suppression will be removed by the KK multiplicity.

7 BRANEWORLDS WITH WARPED EXTRA DIMENSIONS

In this section we describe another way of obtaining 4D gravity on a brane. It is based on a phenomenon of *localization of gravity* discovered by Randall and Sundrum (RS) [10].

We start with a so-called RS II model that has a single brane embedded in a 5-dimensions bulk with negative cosmological constant. The action of the model is written as follows:

$$S_{\rm RS} = \frac{M_*^3}{2}\int d^4x \int_{-\infty}^{+\infty} dy\sqrt{G}(R_5 - 2\Lambda) + \int d^4x\sqrt{g}(T + \mathcal{L}_{\rm SM}(\Psi, M_{\rm SM}))\,, \tag{48}$$

where Λ denotes the negative cosmological constant and T is the brane tension.

The equation of motion derived from this action takes the form (the Gibbons–Hawking surface term in the action is implied and hereafter we put $\mathcal{L}_{\rm SM} = 0$ for simplicity)

$$M_*\sqrt{G}\left(R_{\rm AB} - \frac{1}{2}G_{\rm AB}R\right) = -M_*^3\Lambda\sqrt{G}G_{\rm AB} + T\sqrt{g}\,g_{\mu\nu}\,\delta_A^\mu\,\delta_B^\nu\,\delta(y)\,. \tag{49}$$

In our conventions the brane is located in extra space at the $y = 0$ point. The above equations have a solution with a flat 4D worldvolume

$$ds^2 = e^{-|y|/L}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2\,, \tag{50}$$

where $\eta_{\mu\nu} = \mathrm{diag}(-+++)$ is the four-dimensional flat space metric, and we introduced the following notations

$$L \equiv \sqrt{-\frac{3}{2\Lambda}}\,, \quad T = \frac{3M_*^3}{L}\,. \tag{51}$$

The values of Λ and T have to be carefully adjusted to each other for this solution to exist. Although the coordinate y runs in the interval $(-\infty, +\infty)$ nevertheless, the physical size of extra dimension is finite:

$$\int_{-\infty}^{+\infty} dy\sqrt{G} \sim L\,. \tag{52}$$

The primary question that we would like to address is what does gravity look like on the brane? For this let us consider graviton fluctuations:

$$ds^2 = \left(e^{-|y|/L}\eta_{\mu\nu} + h_{\mu\nu}(x,y)\right)dx^\mu dx^\nu + dy^2\,. \tag{53}$$

We decompose $h_{\mu\nu}(x,y) \equiv u(y)\tilde{h}_{\mu\nu}(x) = u(y)\epsilon_{\mu\nu}\exp(ipx)$ with $p^2 = -m^2$. As a result the equation for the function u takes the following form:

$$\left(-m^2e^{|y|/L} - \partial_y^2 - \frac{2}{L}\delta(y) + \frac{1}{L^2}\right)u(y) = 0\,. \tag{54}$$

For a zero-mode $m^2 = 0$ this equation simplifies and the solution can be found easily:

$$u(y) = \mathrm{const.}e^{-|y|/L}\,. \tag{55}$$

Hence the interval for the zero-mode factorizes as follows

$$ds^2 = e^{-|y|/L}\tilde{g}_{\mu\nu}(x)dx^\mu dx^\nu + dy^2\,, \tag{56}$$

where we used the notations $\tilde{g}_{\mu\nu}(x) \equiv \eta_{\mu\nu} + \tilde{h}_{\mu\nu}(x)$.

It is important to emphasize that the five-dimensional action is integrable w.r.t. y for the zero-mode

$$\frac{M_*^3}{2}\int d^4x\int_{-\infty}^{+\infty} dy\sqrt{G}R_{(5)} \to \frac{M_*^3(2L)}{2}\int d^4x\sqrt{\tilde{g}}\tilde{R}\,. \tag{57}$$

The result of this integration is a conventional 4D action. Hence we find a relation between the 4D Planck mass and M_*

$$M_{\rm Pl}^2 = M_*^3(2L)\,. \tag{58}$$

This looks similar to the relation between the fundamental scale M_*, the size of extra dimension L, and the Planck mass $M_{\rm Pl}$ in the ADD model with one extra dimension. The similarity is due to the fact that the effective size of the extra dimension that is felt by the zero-mode graviton is finite $\sim L$ as in the ADD as well as in the RS models.

Besides the zero-mode there is an infinite number of KK modes [10]. Since the extra dimension is not compactified the KK modes have no mass gap. In the zero-mode approximation used in (57) these

states were neglected. However, at short distances $<< L$, the effects of those modes become important. This can be seen by calculating a static potential between sources on a brane. The result reads:

$$V(r) = -\frac{G_N m_1 m_2}{r}\left(1 + \frac{(2L)^2}{r^2}\right) . \tag{59}$$

The second term in the parenthesis is due to the exchange of KK modes. We see that this term becomes dominant when $r \lesssim L$.

The above construction with the localized graviton can be used for a new solution of the hierarchy problem. This is achieved in a so-called RS I model [30].

The model contains two branes that are placed at the endpoints of an interval of a certain size. One brane, called the 'hidden brane', has positive tension and the other one, called the 'visible brane', has negative tension. The equation of motion for this model looks as follows:

$$M_* \sqrt{G}\left(R_{\rm AB} - \frac{1}{2}G_{\rm AB}R\right) - M_*^3 \Lambda \sqrt{G} G_{\rm AB} =$$
$$T_{\rm hid}\sqrt{g_{\rm hid}}\, g^{\rm hid}_{\mu\nu}\, \delta^\mu_A\, \delta^\nu_B\, \delta(y) \;+\; T_{\rm vis}\sqrt{g_{\rm vis}}\, g^{\rm vis}_{\mu\nu}\, \delta^\mu_A \delta^\nu_B\, \delta(y - y_0)\,, \tag{60}$$

were we used the notations

$$g^{\rm hid}_{\mu\nu}(x) = G_{\mu\nu}(x, y = 0)\,, \quad g^{\rm vis}_{\mu\nu}(x) = G_{\mu\nu}(x, y = y_0)\,. \tag{61}$$

As we mentioned above, the y direction is compactified on an orbifold $S_1/\mathbf{Z}_2$ and y runs in the interval $[-y_0, y_0]$. One can check that there exists the following static solution to the equations of motion

$$ds^2 = e^{-|y|/L}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2\,. \tag{62}$$

The next step is find out fluctuations about this classical background. For this we proceed as in the RS II case. The derivation is straightforward and the result is that the tensor $\eta_{\mu\nu}$ should be replaced as $\eta_{\mu\nu} \to \bar{g}_{\mu\nu}(x)$, where

$$g^{\rm hid}_{\mu\nu}(x) = \bar{g}_{\mu\nu}(x)\,, \quad g^{\rm vis}_{\mu\nu}(x) = e^{-|y_0|/L}\bar{g}_{\mu\nu}(x)\,. \tag{63}$$

Let us now look at what this leads to. For this we turn to the matter part of the Lagrangian. In the RS I case it is assumed that the Standard Model fields are localized on a negative tension brane, i.e., at $y = y_0$. As a representative SM field we consider the Higgs field ϕ. We obtain:

$$\int d^4x \sqrt{g_{\rm vis}}\left\{g^{\mu\nu}_{\rm vis}(D_\mu\phi)^+(D_\nu\phi) - \lambda(|\phi|^2 - v_0^2)^2\right\} \to$$
$$\int d^4x\sqrt{\bar{g}}\left\{\bar{g}^{\mu\nu}(D_\mu\phi)^+(D_\nu\phi) - \lambda(|\phi|^2 - e^{-y_0/L}v_0^2)^2\right\}\,. \tag{64}$$

Hence the Higgs VEV on a visible brane is rescaled by an exponential factor $v = e^{-y_0/2L}v_0$. Thus, all masses on the visible brane are suppressed by this exponential factor as compared to their natural values

$$m^2 = e^{-y_0/L}m_0^2\,. \tag{65}$$

If $m_0 \sim M_{\rm Pl}$, then in order to get $m \sim$ TeV one needs $y_0/L \sim 100$. Therefore, a small hierarchy in y_0/L gives rise to a large hierarchy between m and m_0.

The hierarchy problem is solved at the expense of fine tuning of the tension of the hidden brane to the tension of the visible brane and both these tensions to the bulk cosmological constant. A possible way to avoid the fine tuning is to use the stabilization mechanism proposed by Goldberger and Wise [31]. Another interesting scenario, studied by Karch and Randall [32], emerges when the tension and bulk cosmological constant are slightly detuned so that the worldvolume has AdS_4 geometry. Regretfully, detailed discussion of these developments goes beyond the scope of the present lectures.

Selected topics of RS phenomenology are:

– *Missing energy signals in accelerator experiments.* The SM particles are localized on a brane only up to some energy scale that is comparable with M_*. At about that scale the SM particles could be emitted into the bulk. As in the ADD case, this would provide missing energy signals in accelerator experiments. See Ref. [33] for details.
– *Gauge coupling unification.* In a conventional 4D theory the renormalization group running of the gauge coupling constants is logarithmic. As we discussed before, this changes in flat higher dimensions, the power-law running takes place [20]. However, in the RS case the extra 5th dimension is not flat. This affects dramatically the gauge coupling running which can still be logarithmic as was discussed in Refs. [34], [35].

8 INTRODUCTION TO THE MINIMAL SUPERSYMMETRIC STANDARD MODEL (MSSM)

The MSSM is a minimal supersymmetric extension of the two-Higgs-doublet SM amended with the soft SUSY-breaking terms [36]. Each SM particle is assigned a superpartner. The Higgs sector of the model contains two Higgs doublets and their SUSY partners. The SUSY partners are heavy at the EW scale, but could be observable at a scale between a few hundred GeV and a few TeV that is relevant for the LHC. We shall discuss the basic properties of this model below. For detailed discussion of SUSY and the MSSM phenomenology with references to the original works see, for example, Refs. [37], [38].

8.1 Introduction to supersymmetry

We start with the conventions on spinors in 4D space–time. The Left and Right chirality spinors are defined in a standard way

$$\Psi_L \equiv (1-\gamma_5)\Psi/2\,, \qquad \Psi_R \equiv (1+\gamma_5)\Psi/2\ , \tag{66}$$

where Ψ is a four-component Dirac spinor. These are two-component entities that are also called the Weyl spinors. Yet another useful construct is the Majorana spinor. To discuss the Majorana spinor we define the charge conjugation

$$\Psi^c \equiv C\bar{\Psi}^T\,, \tag{67}$$

where $C \equiv i\gamma^2\gamma^0$. Then, the Majorana spinor is defined as a charge-self-conjugate spinor

$$\Psi^c_M \equiv \Psi_M\,. \tag{68}$$

Not all of the components of the Majorana four-spinor are independent

$$\Psi^T_M = (\Psi^T_L, (-i\sigma_2\Psi^*_L)^T)\,. \tag{69}$$

It is useful to note that any Dirac spinor Ψ gives rise to two Majorana spinors

$$\Psi_{1M} = \frac{i}{\sqrt{2}}(\Psi+\Psi^c) \quad \Psi_{2M} = -\frac{i}{\sqrt{2}}(\Psi-\Psi^c)\,. \tag{70}$$

Having introduced, the basics of the spinors we can take the simplest view of supersymmetry. In SUSY theories in four dimensions the following properties hold:

– For every bosonic degree of freedom (e.g., a complex scalar ϕ) there is a superpartner fermionic degree of freedom (e.g. a Weyl, or Majorana spinor) and vice versa.
– The bosonic and fermionic SUSY partners are degenerate in mass.

An example of a SUSY theory is the supersymmetric version of a theory of a photon. The Lagrangian reads

$$\mathcal{L} = -\frac{1}{4g^2}F_{\mu\nu}^2 + \frac{1}{2}\bar{\lambda}i\hat{\partial}\lambda\,. \tag{71}$$

This describes a massless photon with two physical degrees of freedom and a massless Majorana spin-1/2 state λ which also has two degrees of freedom.

8.2 Introduction to superspace

The Lagrangian (71) is written in terms of the fields of a photon and its SUSY partner photino. However, in complicated SUSY theories it is more useful to unify these two fields (or all superpartners) in a single entity that is called the superfield. We give below an elementary introduction to the superfield formalism.

Let us start with the so-called Grassmann numbers that obey the following properties:

$$\theta_1\theta_2 + \theta_2\theta_1 = 0\,, \tag{72}$$

therefore

$$\theta_1\theta_1 = \theta_2\theta_2 = 0\,. \tag{73}$$

As a result of this property, any analytic function of θ can be expressed as a finite series in powers of θ:

$$f(\theta) = a + b\theta\,. \tag{74}$$

This is just a Taylor series expansion and all the higher order terms are zero because of (73).

As a next step we introduce the Grassmann spinors (two-component anticommuting spinors) which we denote, as customary, by the same letter:

$$\theta_\alpha\theta_\beta + \theta_\beta\theta_\alpha = 0\,, \tag{75}$$

where $\alpha, \beta = 1, 2$. There also exists a conjugate spinor $\bar{\theta}^{\dot{\alpha}}$.

Having introduced these spinors, one could think of an extended space that possesses extra (fermionic) dimensions parametrized by the Grassmann coordinates (this is unlike the ordinary extra dimensions that are described by ordinary numbers)

$$Z \equiv (x_\mu, \bar{\theta}, \theta)\,. \tag{76}$$

These coordinates describe the superspace. Fields defined on this space are called superfields. For the following discussions it is convenient to introduce the notation:

$$y^\mu \equiv x^\mu + i\theta^\alpha\sigma^\mu_{\alpha\dot{\beta}}\bar{\theta}^{\dot{\beta}}\,. \tag{77}$$

Now we are ready to introduce the simplest superfield, the so-called chiral superfield Φ. It contains the following components

$$\begin{aligned}\Phi(y,\theta) = A(y) + \sqrt{2}\theta\chi(y) + \theta^2F(y) = \\ = A(x) + i\theta\sigma^\mu\bar{\theta}\partial_\mu A + \frac{1}{4}\theta^2\bar{\theta}^2\partial^2 A + \sqrt{2}\theta\chi - \frac{i}{\sqrt{2}}\theta^2\partial_\mu\chi\sigma^\mu\bar{\theta} + \theta^2F.\end{aligned} \tag{78}$$

Here A denotes a complex scalar field and χ is a Weyl spinor. Hence, the chiral superfield describes one complex scalar and one Weyl fermion and one auxiliary field F. The superspace action for the chiral superfields can be written as follows:

$$\int d^4x d^2\theta d^2\bar{\theta}\Phi^+\Phi = \int d^4x\left(A^*\partial^2 A + F^*F + \frac{i}{2}\partial_\mu\bar{\chi}\bar{\sigma}^\mu\chi - \frac{i}{2}\bar{\chi}\bar{\sigma}^\mu\partial_\mu\chi\right)\,. \tag{79}$$

This action gives rise to the kinetic terms for the components of the chiral superfield A and χ. As we see, the F component of the chiral superfield is not dynamical since it has no kinetic term, hence it is an auxiliary field.

Using the Chiral superfield one can introduce the very important notion of a superpotential. The superpotential, very roughly, is what replaces an ordinary potential of non-SUSY theories. Very powerful techniques have been developed that allow one to manipulate the superpotentials, hence this notion is extremely useful in model building.

We introduce a superpotential $\mathcal{W}(\Phi)$ as a function that is analytic in Φ. Then we define the interactions as follows:

$$\int d^4x \left(d^2\theta \mathcal{W}(\Phi) + h.c.\right) \equiv \int d^4x \mathcal{L}_W \,. \tag{80}$$

The interaction Lagrangian in terms of components takes the form:

$$\mathcal{L}_W = \sum_i |\frac{\partial \mathcal{W}}{\partial \Phi_i}|^2 - \frac{1}{2}\sum_{i,j}\left(\bar{\chi}_i \frac{\partial^2 \bar{\mathcal{W}}}{\partial \bar{\Phi}_i \partial \bar{\Phi}_j}\bar{\chi}_j + h.c.\right) \,. \tag{81}$$

Hence, the total supersymmetric Lagrangian in superspace takes the form:

$$S_{\rm SUSY} = \int d^4x d^2\theta d^2\bar{\theta}\Phi^+\Phi + \int d^4x \left(d^2\theta \mathcal{W}(\Phi) + d^2\bar{\theta}\bar{\mathcal{W}}(\bar{\Phi})\right) \,. \tag{82}$$

The first term is called the Kähler potential or a D-term. It gives rise to the kinetic terms. The second term is a superpotential. It defines interactions.

Consider as an example the following superpotential:

$$\mathcal{W} = \frac{1}{2}a\Phi^2 + \frac{1}{3}\Phi^3 \,. \tag{83}$$

Then,

$$\frac{\partial \mathcal{W}}{\partial \Phi} = a\Phi + \Phi^2 \,. \tag{84}$$

Therefore, the interaction Lagrangian takes the form

$$\mathcal{L}_W = |a\Phi + \Phi^2|^2 - (\Phi\chi\chi + \Phi^*\bar{\chi}\bar{\chi}) \,. \tag{85}$$

The first term gives a mass $a^2|\Phi|^2$ and the second one is a Yukawa interaction vertex (in all the above expressions we used the conventions $\Psi\chi \equiv \Psi^\alpha\chi_\alpha$ and $\bar{\Psi}\bar{\chi} \equiv \bar{\Psi}_{\dot{\alpha}}\bar{\chi}^{\dot{\alpha}}$).

8.3 Vector superfield

In order to describe gauge interactions, one needs to introduce a superfield that contains vector fields. This is the vector superfield. This can be introduced by looking at the local gauge transformations of a chiral superfield. Under local gauge group a chiral superfield transforms as

$$\Phi(y,\theta) \to e^{i\Lambda(y,\theta)}\,\Phi(y,\theta) \,, \tag{86}$$

where Λ has to be a chiral superfield too. As a result, the chiral kinetic term transforms as follows:

$$\Phi^+\Phi \to \Phi^+\, e^{-i\Lambda^+ + i\Lambda}\,\Phi \,. \tag{87}$$

To make the kinetic term invariant one should introduce a vector superfield V that satisfies the following properties:

$$V^+ \,=\, V \,. \tag{88}$$

Under the gauge transformations the vector superfield transforms as

$$e^{gV} \rightarrow e^{i\Lambda^+} e^{gV} e^{-i\Lambda} . \tag{89}$$

Therefore, the action

$$S = \int d^4x\, d^2\bar{\theta} d^2\theta\, \Phi^+ e^{gV} \Phi , \tag{90}$$

is gauge invariant and supersymmetric.

Let us look at the component form of V. In the so-called Wess–Zumino gauge, V takes a simple form:

$$V = -\theta\sigma^\mu\bar{\theta} v_\mu - i\bar{\theta}^2\theta\lambda + i\theta^2\bar{\theta}^2\bar{\lambda} + \frac{1}{2}\theta^2\bar{\theta}^2 D , \tag{91}$$

where

v_μ is a vector field that describes a spin-1 state;

λ is a Weyl spinor that describes a spin-1/2 state;

D is an auxiliary scalar field.

As a next step we need to find a superfield generalization of the field strength for the vector fields, $F_{\mu\nu}$. In the case of Abelian gauge fields this is achieved by introducing

$$\tilde{W}_\alpha = -\frac{1}{4}\bar{D}\bar{D} D_\alpha V \tag{92}$$

where

$$D_\alpha \equiv \frac{\partial}{\partial\theta^\alpha} + i\,\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\,\partial_\mu . \tag{93}$$

In terms of the component fields we see that W is a SUSY generalization of $F_{\mu\nu}$

$$\tilde{W}_\alpha - i\lambda_\alpha - i\sigma^{\mu\nu\beta}_\alpha\theta_\beta F_{\mu\nu} + \theta_\alpha D + \theta^2\sigma^\mu_{\alpha\dot{\alpha}}\partial_\mu\bar{\lambda}^{\dot{\alpha}}. \tag{94}$$

Thus, a theory of a photon and photino can be written in the following form:

$$S = \int d^4x d^2\theta \left(\frac{1}{4}\tilde{W}_\alpha\tilde{W}^\alpha + \text{h.c.}\right) = \int d^4x \left(-\frac{1}{4}F^2_{\mu\nu} - i\bar{\lambda}\bar{\sigma}^\mu\partial_\mu\lambda + \text{h.c.}\right) \tag{95}$$

Finally, we can also write a most general renormalizable action involving gauge and matter fields

$$S = \int d^4x\, d^2\bar{\theta} d^2\theta\, \Phi^+ e^{gV} \Phi + \left[\int d^4x d^2\theta \left(\frac{1}{4}\tilde{W}_\alpha\tilde{W}^\alpha + \mathcal{W}(\Phi)\right) + \text{h.c.}\right], \tag{96}$$

where $\mathcal{W}$ is a superpotential. In the component form the above action reads

$$\mathcal{L} = -|D_\mu A|^2 - i\bar{\chi}\bar{\sigma}^\mu D_\mu\chi - \frac{1}{4}F^2_{\mu\nu} - i\bar{\lambda}\bar{\sigma}^\mu D_\mu\lambda - i\sqrt{2}\bar{\lambda}\bar{\chi}A + i\sqrt{2}A^*\chi\lambda - \frac{1}{2}\left(\frac{\partial^2\mathcal{W}}{\partial\Phi^2}\chi^2 + \frac{\partial^2\bar{\mathcal{W}}}{\partial\bar{\Phi}^2}\bar{\chi}^2\right) - |\frac{\partial\mathcal{W}}{\partial\Phi}|^2 - \frac{1}{2}g^2(A^*T^{(a)}A)^2 . \tag{97}$$

8.3.1 Particle content of the MSSM

In this section we discuss the particle content of the MSSM. Let us start with the vector superfields in the MSSM. Below we list the superfields, their notation and quantum numbers with respect to the $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge group.

$$\begin{array}{rcc} \text{Gluons and gluions} & V^a & (8,1,0) \\ \text{W}'\text{s and winos}(\tilde{\text{w}}) & V^1 & (1,3,0) \\ \text{B}'\text{s and binos}(\tilde{\text{b}}) & V & (1,1,0) \end{array} \tag{98}$$

The kinetic terms for these particles are written in a standard form:

$$\int d^2\theta \frac{1}{4}\left(\bar{W}^{(a)}\bar{W}^{(a)} + \bar{W}^{(i)}\bar{W}^{(i)} + \bar{W}\bar{W}^{(a)}) + \text{h.c.}\right). \tag{99}$$

The matter fields are described by chiral superfields. A superfield Q contains quark as well as squark doublets

$$(u,d)_L \text{ and } (\tilde{u},\tilde{d})_L\,, \tag{100}$$

note that the subscript L of the squark field is just for notational purposes and does not denote the chirality (squarks are scalars and have no chirality). The superfield $\bar{U}$ contains

$$\bar{u}_R \text{ and } \tilde{u}^*_R\,, \tag{101}$$

and the superfield $\bar{D}$ contains

$$\bar{d}_R \text{ and } \tilde{d}^*_R\,. \tag{102}$$

Finally, the field content of the superfield L is

$$(\nu,e)_L \text{ and } (\tilde{\nu},\tilde{e})_L\,. \tag{103}$$

The assignment of quantum numbers is as follows:

$$\begin{array}{rcc} (u,d)_L \text{ and } (\tilde{u},\tilde{d})_L & Q & (3,2,1/6) \\ \bar{u}_R \text{ and } \tilde{u}^*_R & \bar{U} & (\bar{3},1,-2/3) \\ \bar{d}_R \text{ and } \tilde{d}^*_R & \bar{D} & (\bar{3},1,1/3) \\ (\nu,e)_L \text{ and } (\tilde{\nu},\tilde{e})_L & L & (1,2,-1/2) \\ (e)_R \text{ and } (\tilde{e})^*_R & \bar{E} & (1,1,1) \\ \text{Higgs}(H_1) \text{ and Higgsino}(\tilde{h}_1) & H_1 & (1,2,-1/2) \\ \text{Higgs}(H_2) \text{ and Higgsino}(\tilde{h}_2) & H_2 & (1,2,1/2)\,. \end{array} \tag{104}$$

After spontaneous EW symmetry breaking three Higgs particles are eaten up by $W^\pm$ and Z bosons. What remains is two neutral CP-even bosons h and H^0, two charged bosons $H^\pm$ and one neutral CP-odd boson A. This is unlike the minimal SM where only one neutral Higgs boson exists.

Kinetic terms and interactions for the matter fields can be written as follows:

$$\mathcal{L} = \int d^2\theta d^2\bar{\theta}\,\Phi^+\left(e^{g_1 V + g_2 V^i T^i + g_3 V^a t^a}\right)\Phi\,, \tag{105}$$

where

$$\Phi \equiv (Q,\bar{U},\bar{D},L,\bar{E},H_1,H_2)^T\,, \tag{106}$$

and the coupling constants are

$$e = g_1 \cos\theta_W, \quad \cos\theta_W = \frac{g_2^2}{g_1^2 + g_2^2}, \tag{107}$$

where e denotes the electric charge. The most general form of the renormalizable superpotential is:

$$\begin{aligned} \mathcal{W} = \mu H_1 H_2 + \lambda_U Q \bar{U} H_2 + \lambda_D Q \bar{D} H_1 + \lambda_E L \bar{E} H_1 \\ + \lambda_1 L L \bar{E} + \lambda_2 Q L \bar{D} + \lambda_3 \bar{U} \bar{D} \bar{D} \,. \end{aligned} \tag{108}$$

For the MSSM to be viable λ_1, λ_2 and λ_3 should be suppressed since they give rise to the Baryon (B) and/or Lepton (L) number violation.

8.4 R-parity

R-parity plays an essential role in SUSY phenomenology. This is a discrete Z_2 symmetry under which $\theta \to -\theta$. Scalars and their fermionic partners transform differently under the Z_2

$$(Q, \bar{U}, \bar{D}, L, \bar{E}) \to -(Q, \bar{U}, \bar{D}, L, \bar{E}), \tag{109}$$

while

$$(H_1, H_2) \to (H_1, H_2). \tag{110}$$

Once R-parity is imposed on the MSSM Lagrangian, the terms with λ_1, λ_2 and λ_3 are forbidden by this symmetry. The R-parity of all the SM fields is +1, while the R-parity of the SUSY partners is -1. There are very important consequences of R-parity in the MSSM:

- The number of SUSY particles in a given interaction is always conserved modulo 2. SUSY particles can only be pair-produced from the conventional SM particles.
- The lightest SUSY particle (LSP) should be absolutely stable.
- A generic signal of R-parity-conserving SUSY models is missing energy from non-observed LSP.

  The LSP is widely used in constructing the astrophysical models of dark dark matter.

9 SUSY BREAKING

MSSM is regarded as an effective low-energy theory. SUSY breaking occurs at a high scale and is parametrized by the so-called 'soft' mass terms for scalar members of the chiral multiplets and for the gaugino members of the vector multiplets. The soft terms take the form:

$$\begin{aligned} \mathcal{L}_{\text{soft}} \;=\; & -m_1^2 |H_1|^2 - m_2^2 |H_2|^2 + B\mu (H_1 H_2 + \text{h.c.}) - M_{\tilde{Q}}^2 (\tilde{u}_L^* \tilde{u}_L + \tilde{d}_L^* \tilde{d}_L) \\ & - M_{\tilde{u}}^2 \tilde{u}_R^* \tilde{u}_R - M_{\tilde{d}}^2 \tilde{d}_R^* \tilde{d}_R - M_{\tilde{L}}^2 (\tilde{e}_L^* \tilde{e}_L + \tilde{\nu}_L^* \tilde{\nu}_L) - M_{\tilde{e}}^2 \tilde{e}_R^* \tilde{e}_R - \frac{1}{2} \left(M_3 \bar{\tilde{g}} \tilde{g} + M_2 \bar{\tilde{w}} \tilde{w} + M_1 \bar{\tilde{b}} \tilde{b} \right) \\ & - \frac{g}{\sqrt{2} M_W} \left[\frac{M_d}{\cos\beta} A_d H_1 \tilde{Q} \tilde{d}_R^* + \frac{M_u}{\sin\beta} A_u H_2 \tilde{Q} \tilde{u}_R^* \frac{M_e}{\cos\beta} A_e H_1 \tilde{L} \tilde{r}_R^* + \text{h.c.} \right] \end{aligned} \tag{111}$$

As a consequence of the soft SUSY-breaking terms the mass degeneracy between the SM particles and their SUSY partners is lifted. In particular, the SUSY particles become heavy, typically with masses that are above the EW symmetry-breaking scale but potentially still accessible in soon-to-come LHC experiments. For detailed discussions of the SUSY particle phenomenology the reader is referred to Refs. [37, 38].

ACKNOWLEDGEMENTS

The author is grateful to the organizers of the School for their efforts and warm hospitality.

REFERENCES

[1] A. Sirlin, 'Radiative corrections and the standard model of elementary particles,' arXiv:hep-ph/0312242.

[2] I. Tkachev, Astroparticle physics, in these Proceedings, *2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, 26 August–6 September 2003*; hep-ph/0405168.

[3] H. Georgi, H. R. Quinn and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).

[4] N. Arkani-Hamed, A. G. Cohen and H. Georgi, Phys. Lett. B **513**, 232 (2001) [arXiv:hep-ph/0105239].

[5] G. Gabadadze, 'ICTP lectures on large extra dimensions,' arXiv:hep-ph/0308112.

[6] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Klasse 996 (1921); Reprinted with the English translation in *Modern Kaluza–Klein Theories*, edited by T.W. Appelquist, A. Chodos, P.G.O. Freund (Addison-Wesley, Menlo Park, 1987).

[7] O. Klein, Z. Phys., **37**, 895 (1926); Reprinted with the English translation in *Modern Kaluza–Klein Theories*, edited by T.W. Appelquist, A. Chodos, P.G.O. Freund (Addison-Wesley, Menlo Park, 1987).

[8] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, 'The hierarchy problem and new dimensions at a millimeter,' Phys. Lett. B **429**, 263 (1998) [arXiv:hep-ph/9803315]; N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, 'Phenomenology, astrophysics and cosmology of theories with sub-millimeter dimensions and TeV scale quantum gravity,' Phys. Rev. D **59**, 086004 (1999) [arXiv:hep-ph/9807344].

[9] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, 'New dimensions at a millimeter to a Fermi and superstrings at a TeV,' Phys. Lett. B **436**, 257 (1998) [arXiv:hep-ph/9804398].

[10] L. Randall and R. Sundrum, 'An alternative to compactification,' Phys. Rev. Lett. **83**, 4690 (1999) [arXiv:hep-th/9906064].

[11] G. R. Dvali, G. Gabadadze and M. Porrati, '4D gravity on a brane in 5D Minkowski space,' Phys. Lett. B **485**, 208 (2000) [arXiv:hep-th/0005016].

[12] G. R. Dvali and G. Gabadadze, 'Gravity on a brane in infinite-volume extra space,' Phys. Rev. D **63**, 065007 (2001) [arXiv:hep-th/0008054].

[13] G. Dvali, G. Gabadadze and M. Shifman, 'Diluting cosmological constant in infinite volume extra dimensions,' Phys. Rev. D **67**, 044020 (2003) [arXiv:hep-th/0202174]; G. Dvali, G. Gabadadze and M. Shifman, "Diluting cosmological constant via large distance modification of gravity," arXiv:hep-th/0208096.

[14] C. R. Nappi and L. Witten, "Interacting Lagrangian for massive spin two field," Phys. Rev. D **40**, 1095 (1989).

[15] V. A. Rubakov and M. E. Shaposhnikov, "Do We Live Inside A Domain Wall?," Phys. Lett. B **125** (1983) 136.

[16] K. Akama, "An Early Proposal Of 'Brane World'," Lect. Notes Phys. **176**, 267 (1982) [arXiv:hep-th/0001113].

[17] G. R. Dvali and M. A. Shifman, "Domain walls in strongly coupled theories," Phys. Lett. B **396**, 64 (1997) [Erratum ibid. B **407**, 452 (1997)] [arXiv:hep-th/9612128].

[18] J. Polchinski, "Dirichlet-Branes and Ramond-Ramond Charges," Phys. Rev. Lett. **75**, 4724 (1995) [arXiv:hep-th/9510017]; J. Polchinski, "Lectures on D-branes," in Fields, strings and duality, TASI 96, Boulder, Co, 2–28 June 1996, edited by C. Efthimion and B. Greene (World Scientific, Singapore, 1997), pp. 293–356 [arXiv:hep-th/9611050].

[19] N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali and J. March-Russell, "Neutrino masses from large extra dimensions," Phys. Rev. D **65**, 024032 (2002) [arXiv:hep-ph/9811448];

G. R. Dvali and A. Y. Smirnov, "Probing large extra dimensions with neutrinos," Nucl. Phys. B **563**, 63 (1999) [arXiv:hep-ph/9904211].

[20] T. R. Taylor and G. Veneziano, "Strings And D = 4," Phys. Lett. B **212**, 147 (1988).

[21] K. R. Dienes, E. Dudas and T. Gherghetta, Phys. Lett. B **436**, 55 (1998) [arXiv:hep-ph/9803466]; K. R. Dienes, E. Dudas and T. Gherghetta, "Grand unification at intermediate mass scales through extra dimensions," Nucl. Phys. B **537**, 47 (1999) [arXiv:hep-ph/9806292].

[22] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **544**, 3 (1999) [arXiv:hep-ph/9811291]; E. A. Mirabelli, M. Perelstein and M. E. Peskin, Phys. Rev. Lett. **82**, 2236 (1999) [arXiv:hep-ph/9811337].

[23] S. Cullen and M. Perelstein, "SN1987A constraints on large compact dimensions," Phys. Rev. Lett. **83**, 268 (1999) [arXiv:hep-ph/9903422].

[24] G. R. Dvali and S. H. Tye, "Brane inflation," Phys. Lett. B **450**, 72 (1999) [arXiv:hep-ph/9812483];

[25] G. R. Dvali and G. Gabadadze, "Non-conservation of global charges in the brane universe and baryogenesis," Phys. Lett. B **460**, 47 (1999) [arXiv:hep-ph/9904221]; "Evaporating Global Charges In Brane World," Acta Phys. Polon. B **33** (2002) 2419.

[26] J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, "The ekpyrotic universe: Colliding branes and the origin of the hot big bang," Phys. Rev. D **64**, 123522 (2001) [arXiv:hep-th/0103239]; R. Brandenberger and F. Finelli, "On the spectrum of fluctuations in an effective field theory of the ekpyrotic universe," JHEP **0111**, 056 (2001) [arXiv:hep-th/0109004]; J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, Phys. Rev. D **66**, 046005 (2002) [arXiv:hep-th/0109050].

[27] P. J. Steinhardt and N. Turok, "Cosmic evolution in a cyclic universe," Phys. Rev. D **65**, 126003 (2002) [arXiv:hep-th/0111098]; "A cyclic model of the universe," arXiv:hep-th/0111030.

[28] A. Linde, "Inflationary theory versus ekpyrotic / cyclic scenario," *Talk given at Workshop and Conference on the Future of Theoretical Physics and Cosmology in Honor of Steven Hawking's 60th Birthday, Cambridge, England, 7-10 Jan 2002.* arXiv:hep-th/0205259; R. Kallosh, L. Kofman and A. D. Linde, "Pyrotechnic universe," Phys. Rev. D **64**, 123523 (2001) [arXiv:hep-th/0104073].

[29] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister and S. P. Trivedi, "Towards Inflation in String Theory," arXiv:hep-th/0308055.

[30] L. Randall and R. Sundrum, "A large mass hierarchy from a small extra dimension," Phys. Rev. Lett. **83**, 3370 (1999) [arXiv:hep-ph/9905221].

[31] W. D. Goldberger and M. B. Wise, "Modulus stabilization with bulk fields," Phys. Rev. Lett. **83**, 4922 (1999) [arXiv:hep-ph/9907447].

[32] A. Karch and L. Randall, "Locally localized gravity," JHEP **0105**, 008 (2001) [arXiv:hep-th/0011156].

[33] See, for example, H. Davoudiasl, J. L. Hewett and T. G. Rizzo, JHEP **0304**, 001 (2003) [arXiv:hep-ph/0211377].

[34] A. Pomarol, "Grand unified theories without the desert," Phys. Rev. Lett. **85**, 4004 (2000) [arXiv:hep-ph/0005293].

[35] L. Randall and M. D. Schwartz, "Quantum field theory and unification in AdS5," JHEP **0111**, 003 (2001) [arXiv:hep-th/0108114].

[36] S. Dimopoulos and H. Georgi, Nucl. Phys. B **193**, 150 (1981).

[37] J. A. Bagger, "Weak-scale supersymmetry: Theory and practice," arXiv:hep-ph/9604232.

[38] S. Dawson, "The MSSM and why it works," arXiv:hep-ph/9712464.

QUANTUM CHROMODYNAMICS AND HADRONS: AN ELEMENTARY INTRODUCTION

A. Khodjamirian [1]
Institut für Theoretische Teilchenphysik, Universität Karlsruhe,
D-76128 Karlsruhe, Germany

Abstract
Notes of five lectures given at the 2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, September 2003

1 QUARKS AND GLUONS

1.1 Introduction

In the Standard Model the properties of quarks and leptons are remarkably similar, as far as the electroweak interactions are concerned. The quarks with six different flavours are grouped into the three doublets (u,d),(c,s),(t,b), in a one-to-one correspondence to the three lepton doublets (ν_e,e),(ν_μ,μ), (ν_τ,τ), as shown in the schematic chart in Fig. 1. Quarks and leptons interact in a similar way with the electroweak vector bosons γ, $W^\pm$ and Z. Furthermore, it is anticipated that the universal Higgs mechanism is responsible for the generation of the quark and lepton masses (for more details see Ref. [1]).

In addition, quarks interact 'on their own', revealing their specific property, the colour charge (*colour*), a conserved quantum number which is absent for leptons. A quark of a given flavour has three different colour states with equal masses and electroweak charges. Colour-induced interactions between quarks are mediated by *gluons*, the massless and electroweakly-neutral spin-1 particles.

In these lectures I shall discuss the quark–gluon 'corner' of the Standard Model, introducing the basics of Quantum Chromodynamics (QCD), the theory of quark–gluon interactions. Throughout this survey, the main emphasis will be put on the relation between QCD and *hadrons*, the observed bound states of quarks. This first lecture is devoted to the basic properties of the quark–gluon interactions. I shall frequently refer to Quantum Electrodynamics (QED), the more familiar theory of electromagnetic (e.m.) interactions, which is a useful prototype of QCD.

In Fig. 2 the Feynman diagram of the electron–muon e.m. scattering is drawn together with the analogous diagram of the quark–quark interaction. For definiteness, I specify the quarks as having d and s flavours, the counterparts of e and μ in the Standard Model. The two interactions have many important similarities:

- the colour charged quarks emit and absorb gluons in the same way as the electrically charged leptons emit and absorb photons;
- the gluon and the photon are massless;
- both the gluon and the photon have spin 1.

Being guided by this analogy, one would expect that gluon exchanges generate a Coulomb-type interquark force, similar to the usual attraction/repulsion between the electrically charged particles. In reality, quark–gluon interactions are far more complicated. In particular, since the colour charge has three components, quarks can change their colour states after emitting/absorbing gluons, as indicated on the diagram in Fig. 2. Hence, due to colour conservation, gluons also carry colour quantum numbers, and, as a result, interact with each other. In fact, it is the gluon self-interaction that makes QCD dynamics so peculiar. Yet QCD has one basic property in common with QED. Both theories have specific 'internal' symmetries, named *local gauge symmetries*, to be discussed below.

[1] *After 1.01.04 at Fachbereich Physik, Universität Siegen, D-57068, Siegen, Germany*

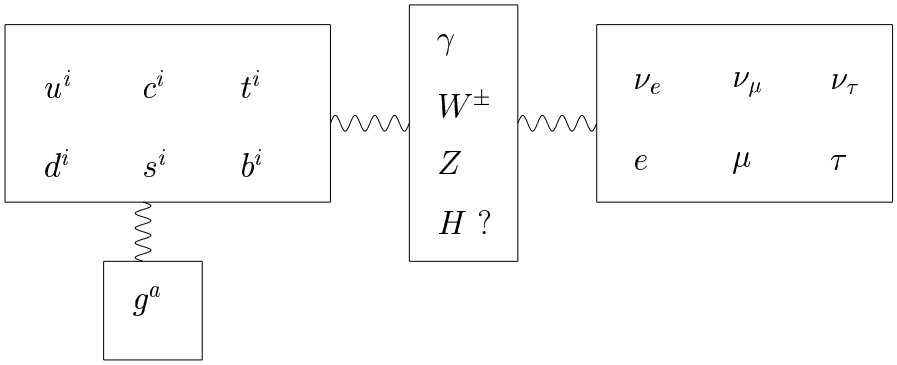


Fig. 1: Particles of the Standard Model; $i = 1, 2, 3$ and $a = 1, ...8$ are the colour indices of quarks and gluons, respectively.

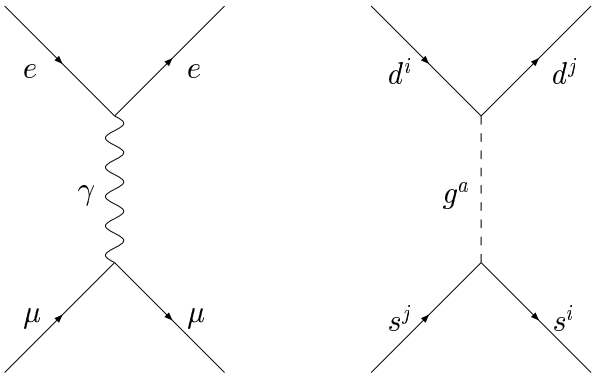


Fig. 2: Diagrams of lepton–photon (left) and quark–gluon (right) interactions. The dashed line is used to denote gluons and the wavy line photons.

1.2 Local gauge symmetry in QED and QCD

The lepton–photon dynamics is described by one compact formula of the QED Lagrangian,

$$L_{QED}(x) = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \sum_{l=e,\mu,\tau} \bar{\psi}_l(x)(iD_\mu\gamma^\mu - m_l)\psi_l(x)\,, \tag{1}$$

where the physical degrees of freedom of the photon field (the electric and magnetic fields) are combined in the field strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, with $\partial_\nu A_\mu \equiv \frac{\partial A_\mu(x)}{\partial x_\nu}$. In the above, $\psi_l(x)$ is the lepton Dirac field with spin 1/2 and mass m_l, and a compact notation for the covariant derivative $D_\mu = \partial_\mu + ieA_\mu$ is introduced. Starting from L_{QED} it is possible to derive Dirac equations for the leptons and Maxwell equations for the photon. Furthermore, Eq. (1) yields the basics elements of the QED Feynman diagrams: the photon and lepton propagators and the lepton–photon interaction vertex. Employing these elements, one obtains the amplitude of a given e.m. process in the form of a perturbative expansion in the numerically small coupling $\alpha_{em} = e^2/4\pi$.

For us the most interesting property of L_{QED} is the gauge symmetry which reveals itself if one locally changes the phase of the lepton fields:

$$\begin{aligned} \psi_l(x) &\rightarrow \psi_l'(x) = exp\,[-i\chi(x)]\,\psi_l(x)\,, \\ \bar{\psi}_l(x) &\rightarrow \bar{\psi}_l'(x) = \bar{\psi}_l(x) exp\,[i\chi(x)]\,, \end{aligned} \tag{2}$$

$\chi(x)$ being an arbitrary function of the 4-coordinate x. If one simultaneously adds the derivative of the same function to the photon field:

$$A_\mu \to A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu \chi(x)\,, \tag{3}$$

the Lagrangian (1) remains invariant. It is easy to check that the transformations (2) and (3) leave intact the physical observables, such as the e.m. current $\bar{\psi}_l \gamma_\mu \psi_l$ or $F_{\mu\nu}$. Importantly, the local gauge symmetry implies that the photon mass vanishes, $m_\gamma = 0$. Indeed, adding to L_{QED} a mass term $m_\gamma^2 A_\mu A^\mu$ for the photon field would evidently violate the symmetry because the latter term changes under (3). The particular case $\chi = \mathrm{const}$ (when only the lepton fields are transformed) corresponds to the *global* gauge symmetry which is responsible for the e.m. current conservation in QED.

From the mathematical point of view the QED gauge transformations form a *group*. Let me remind you that a given set of elements $\{g_i\}$ can be qualified as a group if three conditions are simultaneously fulfilled: 1) a multiplication rule can be defined $g_i * g_k = g_l$, that is, a correspondence is established between a given pair of elements g_i and g_k and a certain third element g_l belonging to the same set; 2) the unit element g_0 exists, so that $g_0 * g_i = g_i$ for each g_i; and 3) the inverse element g_k^{-1} can be specified for each g_k, so that $g_k * g_k^{-1} = g_0$. All three conditions are valid for the (infinite and continuous) set of transformations (2) and (3) generated by the set of the arbitrary functions $\chi(x)$. Indeed, performing two gauge transformations one after the other, with $\chi_1(x)$ and $\chi_2(x)$, is equivalent to the gauge transformation with $\chi_{12}(x) = \chi_1(x) + \chi_2(x)$. The unit element of this 'multiplication' rule is simply the transformation with $\chi_0(x) \equiv 0$ and the inverse element for each $\chi(x)$ is $-\chi(x)$. Importantly, the group multiplication is commutative (the group is Abelian) because the result of the overlap of two phase transformations is independent of their order. The group we are discussing is called $U(1)$. Mathematically, it is equivalent to the group of rotations of the Cartesian coordinate system around one of its axes. The rotation angle plays the same role as the phase χ.

Gauge transformations in QCD have a richer geometry. They are somewhat similar to the general rotations in the 3-dimensional space, involving 3×3 matrices, which do not commute. More specifically, a colour gauge transformation of the quark field

$$\psi_q^i(x) = \begin{pmatrix} \psi_q^1(x) \\ \psi_q^2(x) \\ \psi_q^3(x) \end{pmatrix},$$

with a given flavour $q = u, d, s, ...$, involves transitions between different colour components, a sort of 'rotations of colour coordinates':

$$\begin{aligned} \psi_q^i(x) &\to \psi_q'^i(x) = U_k^i(x)\psi_q^k(x)\,, \\ \bar{\psi}_{q\,i}(x) &\to \bar{\psi}'_{q\,i}(x) = \bar{\psi}_k U_i^{\dagger\,k}(x)\,. \end{aligned} \tag{4}$$

The elements of the 3×3 matrix $U_k^i(x)$ depend arbitrarily on the 4-point x. Furthermore, this matrix is unitary:

$$U_k^{\dagger i} U_j^k = \delta_j^k\,, \tag{5}$$

or, in a symbolic form, $U^\dagger U = 1$. To explain why unitarity is necessary, let me invoke the following physical argument. Since a photon does not distinguish quark colours, the only admissible form for the e.m. interaction of quarks is

$$L_{em}(x) = eQ_q \sum_{k=1,2,3} \bar{\psi}_{q\,k}(x)\gamma_\mu \psi_q^k(x) A^\mu(x)\,, \tag{6}$$

where $Q_q = +2/3\,(-1/3)$ for $q = u, c, t\,(d, s, b)$ and the summation goes over the quark colour indices. The gauge transformation (4) applied to the quark fields yields

$$L_{em}(x) \to L'_{e.m.}(x) = \bar{\psi}_{q\,i}(x) U_k^{\dagger i}(x) \gamma_\mu U(x)_j^k \psi_q^j(x) A^\mu(x)\,. \tag{7}$$

Clearly, only if (5) is valid, does L_{em} remain invariant.

The usual exponential representation of the gauge transformation matrix is:

$$U_k^i(x) = \exp\left[-i \sum_{a=1}^{8} \chi^a(x) \frac{(\lambda^a)_k^i}{2}\right]\,. \tag{8}$$

It contains eight independent and arbitrary functions $\chi^a(x)$ multiplied by eight reference matrices λ^a $(a = 1, ...8)$. The latter have the form chosen by Gell-Mann:

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$
$$\lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\ \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},\ \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{9}$$

The λ-matrices have the following properties: $\lambda^{a\dagger} = \lambda^a$ (hermiticity), $Tr\lambda^a = 0$, and $[\lambda^a, \lambda^b] = f^{abc}\lambda^c$ (noncommutativity), where f^{abc} are totally antisymmetric constants ($f^{123} = -f^{213}$, etc.). It is easy to check that $U(x)$ defined in (8) obeys unitarity and has a unit determinant $\det U = 1$. The (infinite and continuous) set of noncommutative U matrices forms a group, which is called $SU(3)$. One may ask: Why are there eight independent functions χ^a determining the rotations of the three colour states? Isn't eight too many? The point is that the quark fields ψ_q^i are complex functions, hence the matrix $U(x)$ is also a complex function of x determined by 2×9 real functions. Only 8 of them are independent because there are altogether 10 constraints: nine provided by the unitarity relation (5) and one by the unit determinant. With eight 'rotation angles' the group $SU(3)$ is of course quite different from the group of rotations in three dimensions which has only three parameters.

1.3 QCD Lagrangian

The QCD Lagrangian has to be exactly symmetric with respect to the local gauge transformations (4) with matrices (8). This property serves as a guiding principle for constructing L_{QCD}.

We start from the part of the Lagrangian which describes the propagation of free quarks:

$$L_{quark}(x) = \sum_{q=u,d,s,...} \left(\sum_{k=1,2,3} \bar{\psi}_{q\,k}(x)(i\partial_\mu \gamma^\mu - m_q)\psi_q^k(x) \right)\,, \tag{10}$$

and yields the usual Dirac equation for spin 1/2 particle for each quark with a given flavour and colour. Similar to the case of QED, the expression (10) is not invariant with respect to the local gauge transformations (4). An additional term with derivatives of $\chi^a(x)$ remains:

$$L_{quark}(x) \to L_{quark}(x) + \sum_{q=u,d,s,...} \bar{\psi}_{q\,i}(x) \left[i U_k^{\dagger\,i}(x) \partial_\mu U_j^k(x) \right] \gamma^\mu \psi_q^j(x)\,. \tag{11}$$

To restore gauge invariance one follows the same scenario as in QED or in the electroweak theory [1]. The idea (put forward long ago by Yang and Mills) is to introduce 'compensating' spin-1 fields interacting with quark fields. There should be one separate spin-1 field for each of the eight degrees of freedom determining the gauge transformations (4). In this way eight gluons enter the game, with the following quark–gluon interaction term added to the Lagrangian:

$$L_{int}(x) = g_s \sum_{q=u,d,s,...} \bar{\psi}_{q\,i}(x) \frac{(\lambda^a)^i_k}{2} \gamma^\mu \psi^k_q(x) A^a_\mu(x)\,. \tag{12}$$

Here g_s is the dimensionless coupling analogous to e in L_{QED}. In contrast to e.m. interaction, where the photon field is electrically neutral, the gluon fields also carry colour charge, so that the colour state of a quark changes after emitting/absorbing a gluon. The colour of the gluon distinguished by the index $a = 1...8$ can be identified with a superposition of quark and antiquark colours. For example, the gluon field $A^1_\mu(x)$ is in the same colour state as the quark–antiquark pair $\bar{\psi}_{q\,1}\psi^2_q + \bar{\psi}_{q\,2}\psi^1_q$. The A^a_μ-fields in (12) have to be gauge-transformed in the following way:

$$\frac{\lambda^a}{2} A^a_\mu(x) \to U(x) \frac{\lambda^a}{2} A^a_\mu(x) U^\dagger(x) - \frac{i}{g_s} \partial_\mu U(x) U^\dagger(x)\,, \tag{13}$$

so that the overall change of L_{int} cancels the symmetry breaking term on the r.h.s. of (11). It is a simple exercise to check that the combination of transformations (4) and (13) leaves the sum $L_{quark} + L_{int}$ invariant. We see that in QCD the 'compensating' transformation of gluon fields (13) is more complicated than its analog (3) in QED: the addition of the derivative over χ-functions is accompanied by a 'colour rotation'.

To complete the Lagrangian one has to add a gauge-invariant term describing the propagation of gluon fields:

$$L_{glue}(x) = -\frac{1}{4} G^a_{\mu\nu}(x) G^{a\,\mu\nu}(x)\,, \tag{14}$$

where

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu\,, \tag{15}$$

is the gluon field-strength tensor. The local gauge invariance of L_{glue} implies that gluons are massless. At the same time, $G^a_{\mu\nu}$ is a more complicated object than its QED analog $F_{\mu\nu}$. Indeed, substituting (15) to (14), we notice that not only the terms quadratic in A^a_μ emerge (propagation of gluons) but also the three- and four-gluon vertices (gluon self-interactions). Note that, formally, both properties of gluons, colour quantum number and self-interactions, are due to the noncommutativity of the gauge-transformation group (the fact that $f_{abc} \neq 0$). The final form of the QCD Lagrangian is obtained by adding together the three pieces introduced above:

$$\begin{aligned} L_{QCD} &= L_{glue} + L_{quark} + L_{int} \\ &= -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \sum_q \bar{\psi}_q (i D_\mu \gamma^\mu - m_q) \psi_q\,, \end{aligned} \tag{16}$$

where $D_\mu = \partial_\mu - i g_s \frac{\lambda^a}{2} A^a_\mu$. To summarize, L_{QCD} describes not only quark–gluon interactions but also *gluodynamics*, the specific gluon self-interactions which have no analog in QED [1].

[1] In QCD 'light emits light' at the level of the fundamental interactions entering Lagrangian, as opposed to QED where light-by-light interaction appears only as $O(\alpha^2_{em})$ quantum correction (when photons exchange virtual electrons via box diagrams). For a classical Maxwellian electrodynamics, self-interacting e.m. fields would mean 'new physics beyond standard theory'. I am not aware of any discussions of photon self-interactions in the times before quantum field theory. Interestingly, the light emitting light was mentioned in poetry. I found the following sentence written in the XIIIth century by the Armenian poet Kostandin Erznkazi [2]: "And so the light was born from the light, the great light of sun..." (in translation from Armenian).

Table 1: Propagators and vertices in QCD

Quark propagator	$\langle 0\|\psi_q^i(x)\bar{\psi}_{qk}(0)\|0\rangle = i\delta_k^i \int d^4p \left(\frac{p_\alpha\gamma^\alpha + m_q}{p^2 - m_q^2}\right) e^{-ipx}$
Gluon propagator	$\langle 0\|A_\mu^a(x)A_\nu^b(0)\|0\rangle = -i\delta^{ab} \int d^4k \frac{g_{\mu\nu}}{k^2} e^{ikx}$
Quark–gluon vertex	$g_s\bar{\psi}_q(x)\gamma_\mu\frac{\lambda^a}{2}\psi_q(x)A^\mu(x)$
3-gluon vertex	$-\frac{g_s}{2}f^{abc}[\partial_\mu A_\nu^a(x) - \partial_\nu A_\mu^a(x)]A^{b\mu}(x)A^{c\nu}(x)$
4-gluon vertex	$-\frac{g_s^2}{4}f^{abc}f_{ade}A_\mu^b(x)A_\nu^c(x)A^{d\mu}(x)A^{e\nu}(x)$

The quark and gluon propagators and vertices derived from L_{QCD} are collected in Table 1. The formula of the gluon propagator has a certain degree of freedom related to the fact that the physical massless gluon has only two polarization/spin states whereas the field A_μ^a has four components. To make things work, one follows the same recipe as in QED. An additional constraint on the gluon field is introduced, the so called gauge-fixing condition. The gluon propagator given in Table 1 corresponds to the usual Feynman gauge. Note that in QCD purely gluonic loop diagrams are possible, in which case one has to take care of subtracting the contributions of unphysical components of A_μ^a also in these loops. It is usually done by adding specially designed fictitious particles (the so called Fadeev–Popov ghosts) which only appear in the loops, and are not shown in Table 1.

Feynman diagrams in QCD are obtained by employing the quark–gluon propagators and vertices as building blocks. However, the use of diagrams makes sense only if the perturbative expansion in g_s is meaningful. To obey this condition, the *quark–gluon coupling*

$$\alpha_s = \frac{g_s^2}{4\pi}\,, \tag{17}$$

the QCD analog of the e.m. coupling $\alpha_{em} = e^2/4\pi$, has to be sufficiently small, $\alpha_s \ll 1$. If this condition is fulfilled, then, for example, the $O(\alpha_s)$ diagram of the quark–quark interaction in Fig. 2 dominates over the higher-order diagrams with additional gluon exchanges. One has then a tractable situation, with quarks and gluons propagating quasi-freely. Thus, there is an important question to be addressed: How large is α_s?

1.4 Running of the quark–gluon coupling

To answer the above question, one has to investigate quantum effects in QCD, i.e., creation and annihilation of virtual gluons and quarks described by loop diagrams. In QED, which we use as a prototype, quantum loops generate very small effects, such as Lamb shift (the correction to the Coulomb force due to the virtual electron–positron pairs) or the $O(\alpha_{em})$ one-loop correction to the muon magnetic moment. These effects, being accessible in precision experiments, play a minor role in the bulk of electromagnetic processes.

In QCD, quark–gluon loops are far more pronounced and play a crucial role in determining α_s. To have a closer look, let us consider the quark–quark scattering amplitude. In the lowest-order (at the tree

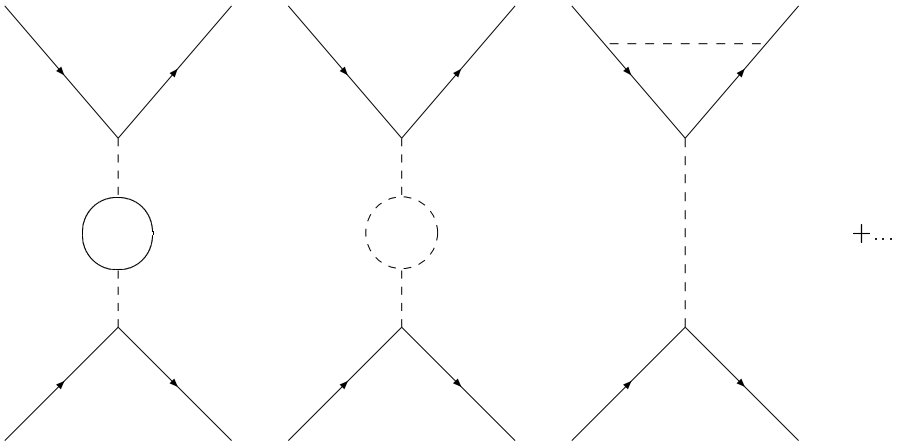


Fig. 3: Some of the diagrams corresponding to quantum loop corrections to the quark–quark scattering.

level) the amplitude is given by the one-gluon exchange diagram shown in Fig. 2 for the particular choice of quark flavours. Having at hand the Feynman rules in QCD it is easy to write down the amplitude:

$$\mathcal{A} = \frac{4\pi\alpha_s}{q^2}\left(\bar{\psi}_{s\,i}\gamma_\mu\frac{(\lambda^a)^i_k}{2}\psi^k_s\right)\left(\bar{\psi}_{d\,j}\gamma^\mu\frac{(\lambda^a)^j_l}{2}\psi^l_d\right)\,, \tag{18}$$

where $q^2 < 0$ is the momentum transfer squared. Below I shall also use the notation for the momentum scale: $Q \equiv \sqrt{Q^2}$, where $Q^2 = -q^2 > 0$. Considering $O(\alpha_s)$ corrections to the amplitude (18) one encounters the diagrams shown in Fig. 3. They contain gluon or quark loops inserted within the exchanged gluon line or in the vertices. These loop effects turn out to be extremely important for evaluating α_s.

Let me first explain the loop diagram calculation in more detail. After substituting propagators of virtual particles, one arrives at four-dimensional Feynman integrals over the 4-momentum k flowing inside the loop, typically:

$$I_{loop}(q, m_q) = \int \frac{d^4k}{(2\pi)^4}\frac{...}{(k^2 - m_q^2)((k+q)^2 - m_q^2)}\,, \tag{19}$$

where the explicit expression in the numerator (depending on the spin of the loop particles) does not play a role in our discussion. For simplicity, I shall also put to zero the quark masses, a reasonable approximation if $Q \gg m_q$. To calculate loop integrals, one usually employs the method of dimensional regularization. The idea is to lower the number of dimensions in the integral (19) making it convergent. One replaces 4 by a generic integer number D, then calculates the integral as a function of D and after that considers the result at an arbitrary noninteger $D = 4 - \epsilon$, schematically:

$$I_{loop}(q, 0) = \int d^4k f(k, q) \to \mu^{(4-D)}\int d^Dk f(k, q) = I(q, \mu, D) \to I(q, \mu, 4-\epsilon)\,. \tag{20}$$

The auxiliary mass scale μ is introduced to keep unchanged the physical dimension of the integral. The major advantage of dimensional regularization is in preserving the gauge symmetry of the amplitudes at each step of the calculation. The divergent part of the integral in this scheme is represented in the form of terms proportional to $1/\epsilon$. A generic expression for the loop integral has the form:

$$I_{loop}(q, 0) \to \mu^{4-D}\int \frac{d^Dk}{(2\pi)^D}\frac{...}{k^2(k+q)^2} = I_1\log(-q^2/\mu^2) + \frac{1}{\epsilon} + I_2\,, \tag{21}$$

where $I_{1,2}$ are calculable finite coefficients.

QCD is a *renormalizable* theory (similar to QED), which means one can absorb all divergent $1/\epsilon$ terms into the so-called Z-factors. Multiplying by Z_g^{-1} the coupling g_s, and by the corresponding factors the quark masses, quark and gluon fields, one defines the finite (renormalized) quantities; e.g., the renormalized coupling is $g_s^{ren} = Z_g^{-1} g_s$. Having in mind the validity of the renormalization procedure I shall simply ignore divergent terms appearing in (21) and in other loop integrals.

Adding the $O(\alpha_s)$ diagrams in Fig. 3 to the tree-level amplitude results in the same expression (18), but with α_s replaced by an *effective coupling* depending on the momentum scale:

$$\alpha_s^{eff}(Q) \equiv \alpha_s \left[1 - \frac{\alpha_s}{4\pi}\left(\beta_0 \log\frac{Q^2}{\mu^2} + const\right)\right], \tag{22}$$

where a shorthand notation

$$\beta_0 = 11 - \frac{2}{3} n_f \tag{23}$$

is introduced. Here n_f is the number of 'active' quark flavours in the loop diagrams. Only those quarks which have $m_q \ll Q$ contribute to $\alpha_s^{eff}(Q^2)$. Importantly, β_0 is positive, because the term 11 originating from the gluon loops exceeds the quark-loop contribution $-2n_f/3$ (since $n_f < 6$ in any case).

Taking α_s^{eff} at a different scale Q_0,

$$\alpha_s^{eff}(Q_0) \equiv \alpha_s \left[1 - \frac{\alpha_s}{4\pi}\left(\beta_0 \log\frac{Q_0^2}{\mu^2} + \text{const}\right)\right], \tag{24}$$

and dividing (22) by (24) one obtains, with an accuracy of $O(\alpha_s^2)$:

$$\alpha_s^{eff}(Q) = \alpha_s^{eff}(Q_0)\left[1 - \frac{\alpha_s^{eff}(Q_0)}{4\pi}\beta_0 \log\frac{Q^2}{Q_0^2}\right], \tag{25}$$

the relation between effective couplings at two different scales. The approximation (25) is valid only if $\alpha_s^{eff}(Q_0)$ is sufficiently small and the higher-order corrections are negligible. Remarkably, (25) predicts that $\alpha_s^{eff}(Q)$ decreases when the momentum scale Q increases. Thus, if the perturbative expansion in α_s is applicable at certain Q_0, it should behave even better at $Q > Q_0$.

Note that (25) still has to be improved. At $Q \to \infty$ the logarithm becomes very large driving the combination $\alpha_s \log Q^2$ to rather big values, so that the $O\left(\left[\alpha_s \log\left(Q^2/Q_0^2\right]^2\right)\right.$ correction originating from the two-loop diagrams shown in Fig. 4 becomes important, and the whole perturbative construction is again in danger. Fortunately, a systematic resummation of all $O\left(\left[\alpha_s \log\left(Q^2/Q_0^2\right]^n\right)\right.$ corrections is possible. In practice, one does not need to calculate all multiloop diagrams, which would be a tremendous work. Instead, the renormalization-group method is used, which is, however, beyond our scope. The resulting expression for the *running coupling* is:

$$\alpha_s(Q) = \frac{\alpha_s(Q_0)}{1 + \frac{\alpha_s(Q_0)}{4\pi}\beta_0 \,.\, \log\frac{Q^2}{Q_0^2}} \tag{26}$$

(hereafter the superscript 'eff' is omitted). Expanding the denominator in (26) and retaining only the first two terms we return to the relation (25).

1.5 Asymptotic freedom

A consistent use of the running QCD coupling is possible if one can find a range of Q where $\alpha_s(Q)$ is numerically small. The first indications that α_s is indeed small at large momentum transfers were obtained in the studies of deep-inelastic lepton–nucleon scattering (to be discussed in Lecture 3). This remarkable discovery paved the way for using QCD perturbation theory with the running coupling in many other processes.

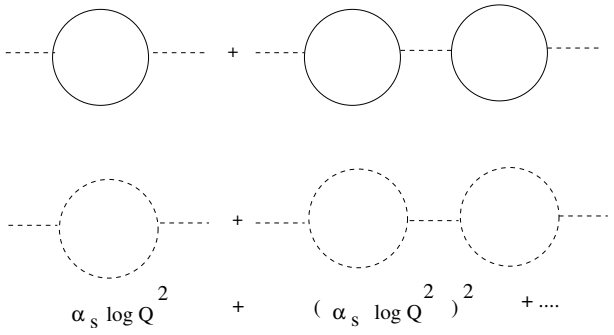


Fig. 4: One- and two-loop diagrams contributing to the running of α_s.

To trace the numerical behaviour of $\alpha_s(Q)$ one has to fix the coupling at a certain large scale using an experimental input. One possibility is the decay of Z boson to a quark–antiquark pair. Quarks originating in this decay inevitably build hadrons in the final state (see the next subsection). To avoid hadronic complexity, one measures the total inclusive width $\Gamma(Z \to hadrons)$, so that the probabilities of all possible quark $\to$ hadron transitions add up to a unit. The majority of $Z \to hadrons$ events observed at LEP has a spectacular structure of two hadronic jets originating from the initial, very energetic quark and antiquark ($E_q = E_{\bar{q}} = m_Z/2$ in the Z rest-frame, see, for example, Ref. [3]). On the other hand, the share of $\geq$ 3-jet events in $Z \to hadrons$, with additional jets originating from gluons and/or from secondary quark–antiquark pairs, is small. This observation clearly indicates that the quark–gluon coupling at the scale m_Z is small, or in other words the initial quark pair interacts weakly during the short time after its creation.

The perturbative diagrams of $Z \to q\bar{q}$ ($q = u, d, s, c, b$) including the gluon emission $Z \to \bar{q}qG$ are shown in Fig. 5. Evaluating these diagrams one gets the perturbative expansion for the total hadronic width, schematically:

$$\Gamma(Z \to hadrons) = \sum_{q=u,d,s,..} \Big[\Gamma(Z \to \bar{q}q) \left(1 + C_1^q \alpha_s(m_Z) + C_2^q \alpha_s^2(m_Z) + ...\right) + \Gamma(Z \to \bar{q}qG) \left(1 + C_1^{qG} \alpha_s(m_Z) + ...\right) + ... \Big] , \tag{27}$$

where $\Gamma(Z \to \bar{q}q)$ and $\Gamma(Z \to \bar{q}qG)$ (the latter starting from $O(\alpha_s)$) are the corresponding perturbative widths, and $C_{1,2}^q, C_1^{qG}, ...$ are the calculable coefficients. Smallness of α_s allows one to neglect all higher-order corrections starting, say from $O(\alpha_s^3)$. In the above expression α_s is taken at the scale m_Z, the characteristic scale in this process. One can trace how the running coupling builds up in (27) by summing up all logarithmic corrections generated by the loop insertions similar to the one shown in Fig. 5(c). Comparing the result (27) of the theoretical calculation with the experimental data yields [4]:

$$\alpha_s(m_Z) \simeq 0.12 \ . \tag{28}$$

As expected, it is a rather small number, so that the whole perturbative treatment turns out to be quite consistent. Using (26) and (28), one predicts $\alpha_s(Q)$ at lower scales. The curve plotted in Fig. 6 is taken from Ref. [4] and reflects the current status of the running coupling, including all known (and reasonably small) higher-order corrections to (26). As can be seen in this figure, there is a wide region spreading up to $Q \sim 1$ GeV, where α_s is small and perturbative QCD is applicable. Furthermore, α_s extracted from various processes at different scales agrees with the running behaviour predicted in QCD. The most spectacular consequence of (26) is the vanishing of the running quark–gluon coupling at $Q \to \infty$, revealing that QCD is asymptotically free.

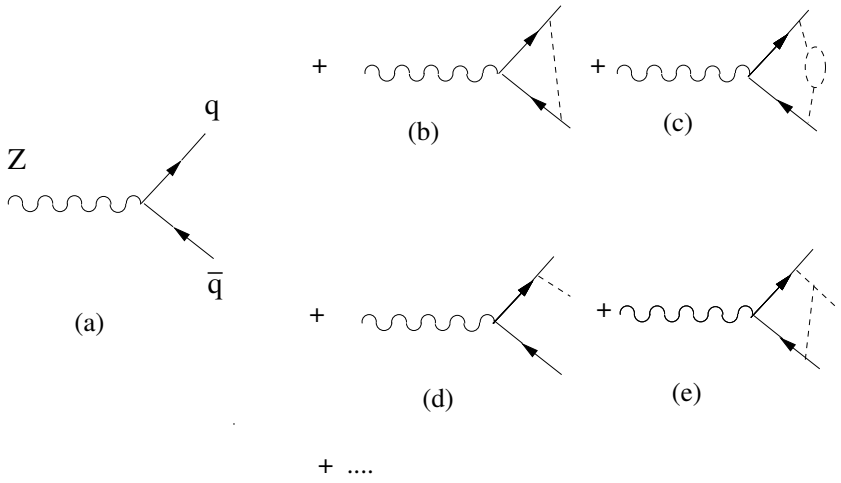


Fig. 5: The lowest-order diagram (a) and some of the higher-order in α_s diagrams (b)–(d) determining the total width Z $\to$ hadrons.

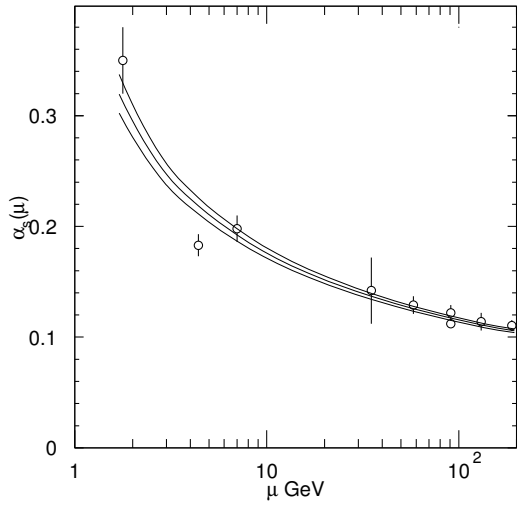


Fig. 6: The dependence of $\alpha_s(\mu)$ on the running scale μ [4]: the upper and lower curves indicate the theoretical uncertainty, the points with errors are the values of α_s extracted at different scales using various methods.

1.6 Confinement and hadrons

Quite an opposite situation takes place in the quark–gluon interactions at small momentum transfers (at long distances). According to (26), if one starts from $Q \gg 1$ GeV and drifts towards smaller scales, α_s grows to $O(1)$ at $Q < 1$ GeV (see Fig. 7). Perturbation theory becomes useless in this region, because in the quark-quark scattering, for example, an infinite amount of higher-order quark–gluon diagrams has to be taken into account. Moreover, at a certain momentum scale denoted by Λ_{QCD} the denominator in (26) vanishes and $\alpha_s(Q)$ diverges. The relation between $\alpha_s(Q)$ and Λ_{QCD} following from (26) is:

$$\alpha_s(Q) = \frac{2\pi}{\beta_0 \log\left(\frac{Q}{\Lambda_{QCD}}\right)}\,. \tag{29}$$

The experimentally measured value (28) corresponds, roughly, to

$$\Lambda_{QCD} = 200{-}300 \text{ MeV}. \tag{30}$$

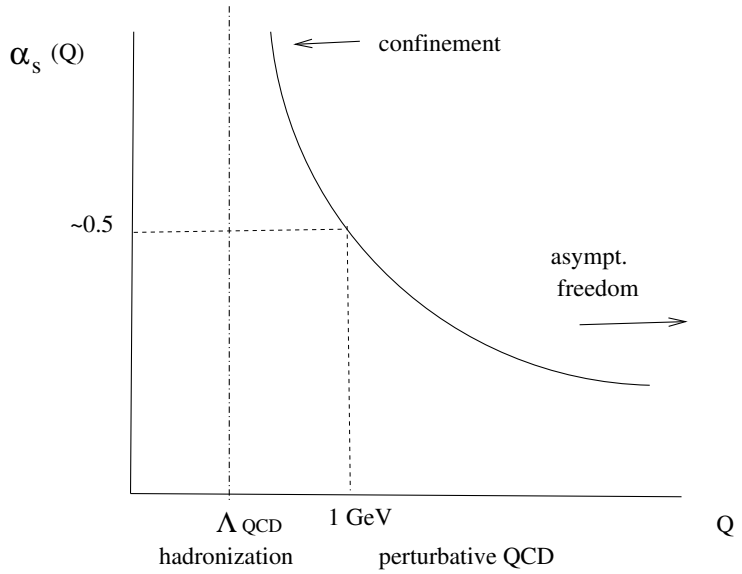


Fig. 7: Schematic view of the α_s behaviour at different scales.

I skip some details concerning the dependence of Λ_{QCD} on the number of active quark flavours, and on the theoretical scheme of the loop-diagram evaluation, relevant for the higher-order corrections to (29). The important fact is that (29) and (30) were derived neglecting the quark masses. We could have put to zero the quark masses in L_{QCD} from the very beginning, starting with a theory without dimensionful parameters. The emergence of an intrinsic energy scale in a theory with dimensionless coupling g_s (*dimensional transmutation*) is a specific property of QCD, due entirely to quantum effects.

The breakdown of perturbation theory and the exploding behaviour of $\alpha_s(Q)$ at $Q \to \Lambda_{QCD}$ are actually anticipated. Long before QCD was invented it was known that at long distances quarks and antiquarks strongly interact and form *hadrons*, the quark–antiquark (*meson*) and 3-quark (*baryon*) bound states. The properties of hadrons will be considered in a more systematic way in the next Lecture. For the present discussion it is important that the characteristic energy scale of hadronic interactions is of $O(\Lambda_{QCD})$. Hence, it is quite natural that the formation of hadrons is due to the strong, nonperturbative quark–gluon force emerging in QCD at momenta $\sim \Lambda_{QCD}$.

Moreover, hadronic matter is the only observable form of quarks and gluons at long distances. In any process of quark and antiquark production, independent of the energy/momenta involved, quarks form hadrons in the final state[2]. Note that in QED the situation is quite different: isolated leptons and other electrically charged particles are observed, and the e.m. bound states (e.g., hydrogen atom, positronium or muonium) can always be split into constituents if a sufficient energy is supplied.

In QCD, the non-observation of free colour-charged particles (quarks, antiquarks and gluons) is arranged in a form of the *colour confinement* principle, postulating that all observable states, i.e., all hadrons, have to be colour-neutral. The confinement principle was never rigorously proved, because of our limited ability to work with QCD beyond perturbation theory. Nevertheless, all experimental results concerning hadrons, as well as lattice simulations of QCD at long distances, unambiguously support colour confinement.

Before the era of QCD the search for free quarks was very popular among experimentalists. The fractional electric charge was a smoking-gun signal to be observed. The hunt for quarks was a part of many accelerator experiments. Not surprisingly, free quarks have never been found at accelerators or in other places (in cosmic rays, water, ice, meteorites etc.). Nowadays, one would hardly invest efforts in the search for free quarks. We are confident that QCD obeys confinement.

[2]There is one exception: the t-quark, decaying via weak interactions, is too short-lived to be bound by quark–gluon forces.

(a) (b)

Fig. 8: The free quark propagator (a) and the gluon correction to it (b).

1.7 Quark masses

The current intervals of quark masses presented in the Particle Data Tables [4] are:

$$\begin{aligned} m_u = 1.5 \div 4.5 \text{ MeV}, \quad m_c = 1.0 \div 1.4 \text{ GeV}, \quad m_t = 174.3 \pm 5.1 \text{ GeV}; \\ m_d = 5 \div 8.5 \text{ MeV}, \quad m_s = 80 \div 155 \text{ MeV}, \quad m_b = 4.0 \div 4.5 \text{ GeV}\,. \end{aligned} \tag{31}$$

The fact that the masses are spread over five orders of magnitude, is a reflection of some fundamental flavour physics not related to QCD, e.g., the Higgs mechanism of the Standard Model. Hence, the masses m_q entering QCD Lagrangian (16) are 'external' parameters. At the same time quark masses are evidently changed in the presence of quark–gluon interactions. At long distances, within hadrons, each quark acquires, roughly speaking, an extra addition of $O(\Lambda_{QCD})$ to its 'bare' mass. It is, however, very difficult, if not impossible, to define this *constituent* quark mass in a model-independent way. Representing the hadron mass as a sum of the constituent quark masses plus some interaction energy, e.g., for a meson:

$$M_{meson} = m_q^{constit} + m_{\bar{q}}^{constit} + E_{int}\,, \tag{32}$$

one can always redistribute the part of E_{int} between the quark and antiquark masses. Since free, on-shell quarks are not observed, the usual definition of the particle mass (the minimal possible energy of the one-particle state) does not work.

The mass values presented in (31) have nevertheless quite a definite meaning. These are 'short-distance' masses of the virtual quarks. As we already know, at large virtualities quark propagation is quasi-free and a consistent use of perturbation theory derived from QCD Lagranigian is possible. In particular, the free quark propagator

$$S(p) = \frac{p_\alpha \gamma^\alpha + m_q}{p^2 - m_q^2} \tag{33}$$

is applicable at $|p^2| \gg \Lambda^2_{QCD}$, with the bare mass m_q from L_{QCD}. Furthermore, at short distances the quark–gluon interactions are calculable in terms of series in α_s. The quark propagating at short distances can emit and absorb a gluon (see diagram (b) in Fig. 8). From the experience with the running coupling, one expects such loops to be important. Adding the loop diagram to the free propagator yields the same expression as (33) with m_q replaced by an effective mass depending on the momentum scale

$$m_q^{eff}(Q) = m_q \left[1 - \frac{\alpha_s}{4\pi}\left(\gamma_0 \log\frac{Q^2}{\mu^2} + \text{const}\right)\right]\,, \tag{34}$$

where I omit the divergent part knowing that it will be absorbed by renormalization. In this case the scale $Q \sim \sqrt{|p^2|}$ is determined by the virtuality of the quark and $\gamma_0 = 4$ is the result of the explicit calculation. Again, as in the case of α_s, we can relate the effective masses at two different scales. Writing down the above equation for another scale Q_0 and using instead of m_q^{eff} a conventional notation $\overline{m}_q$ we obtain:

$$\overline{m}_q(Q) = \overline{m}_q(Q_0)\left(1 - \left(\frac{\gamma_0}{\beta_0}\right)\frac{\alpha_s(Q_0)}{4\pi}\beta_0 \log\frac{Q^2}{Q_0^2}\right)\,, \tag{35}$$

where I have multiplied and divided the logarithmic term by β_0 and used $\alpha_s = \alpha_s(Q_0)$ which is correct with $O(\alpha_s)$ accuracy. With the same accuracy, the expression in parentheses can be transformed further:

$$1 - \left(\frac{\gamma_0}{\beta_0}\right) \frac{\alpha_s(Q_0)}{4\pi} \beta_0 \log \frac{Q^2}{Q_0^2} \simeq \left(1 - \frac{\alpha_s(Q_0)}{4\pi} \beta_0 \log \frac{Q^2}{Q_0^2}\right)^{\gamma_0/\beta_0} . \tag{36}$$

Using (25) we obtain

$$\bar{m}_q(Q) = \bar{m}_q(Q_0) \left(\frac{\alpha_s(Q)}{\alpha_s(Q_0)}\right)^{\gamma_0/\beta_0} , \tag{37}$$

the formula for the *running mass*. A more rigorous derivation is possible using the renormalization group method. Also in recent years, the higher-order corrections to (37) have been calculated.

The masses presented in Ref. [4] are the running masses (also called $\overline{MS}$ masses if one specifies the appropriate renormalization procedure) normalized at some large scale. The light u, d, s quark masses are traditionally taken at $Q = 2$ GeV, whereas a more appropriate scale for the heavy c and b quark masses is the quark mass itself, $Q = m_c$ and $Q = m_b$, respectively, (which means, for example, the virtuality of the c quark is $p^2 = -m_c^2$). The fact that quark masses run with the scale is in accordance with the absence of isolated quarks among observable states.

1.8 Two branches of QCD

To summarize, QCD yields two qualitatively different pictures of quark–gluon interactions:

1) at high-momentum transfers, i.e., at short average distances, perturbative expansions in α_s are applicable in terms of Feynman diagrams with quark and gluon propagators and vertices. In this region the scale-dependence (running) of the coupling and quark masses should be properly taken into account.

2) at low scales, that is, at long distances, one loses control over perturbative interactions between individual quarks and gluons; the latter strongly interact and form hadrons.

Accordingly, QCD is being developed in two different directions. The first one deals with short-distance physics accessible at high energies. One studies specific processes/observables calculable (at least partly) in a form of a perturbative expansion in α_s. A typical short-distance process is the jet production in Z decays considered above, other examples will be presented in Lecture 3.

The second direction deals with nonperturbative quark–gluon interactions at long distances and with hadron dynamics. A complete analytical evaluation of hadronic masses and other parameters directly from L_{QCD} is not yet accessible. Instead, a powerful numerical method of simulating QCD on the space–time lattice has been developed. Lattice QCD has become a separate field, which is beyond the scope of these lectures (for a pedagogical introduction see, for example, Ref. [5]). Still there are a lot of interesting advances in the long-distance 'branch' of QCD, so I shall only be able to cover a part of them. As demonstrated in Lecture 2, many important features of hadron spectroscopy follow from QCD at the qualitative level. The relation of long-distance dynamics to the nontrivial structure of the QCD vacuum will be discussed in Lecture 4. An approximate analytical method of QCD sum rules based on this relation and used to calculate hadronic parameters will be overviewed in Lecture 5.

One might think that physics of hadrons plays a secondary role, because the most important direct tests of QCD in terms of quarks and gluons are done at short distances. Let me emphasize the fundamental importance of hadron dynamics by mentioning two topical problems:

1) *The origin of the nucleon mass*

The proton and the neutron are the lowest and most stable baryons, with the quark content uud and udd, respectively. Their masses

$$m_p \simeq m_n \simeq 940 \text{ MeV} , \tag{38}$$

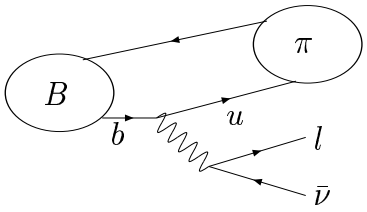


Fig. 9: $B \to \pi l \nu_l$ decay involves a hadronic transition between B and π states.

are substantially larger than the tripled quark mass $m_{u,d} = O(\text{ few MeV})$. We conclude that $\sim 99\%$ of the baryon matter in the Universe is not related to the 'fundamental' quark masses generated by the Higgs mechanism or from some other flavour dynamics. The bulk of the baryon mass is due to the long-distance quark–gluon interactions. Indeed, adding a 'constituent' mass of $O(\Lambda_{QCD})$ to each quark, one gets the order-of-magnitude value of $m_{p,n}$. Certainly, the problem of the nucleon mass is quite fundamental and has to be solved within long-distance QCD.

2) *Extracting electroweak parameters from B decays*

The weak decays of B mesons (the bound states of b quark and light antiquark) represent a valuable source of information on fundamental aspects of electroweak interactions, such as the quark-mixing CKM matrix and the origin of CP violation (see Ref. [6]). One topical example is the $B \to \pi l \nu_l$ decay observed at B factories [4]. This decay (see Fig. 9) is driven by the $b \to u$ weak transition, proportional to V_{ub}, one of the poorly known CKM matrix elements. In order to extract this fundamental parameter from the experimentally measured partial width, one needs to divide out the hadronic $B \to \pi$ transition amplitude (form factor). The latter is essentially determined by the long-distance quark–gluon interactions.

Below, in Lecture 5 we shall discuss the (approximate) solutions of the two above-mentioned problems.

2 FROM QUARKS TO HADRONS

2.1 Mesons and baryons

Let us now have a closer look at the properties of hadrons[3], the bound states of quarks. Much experimental data on hadrons is accumulated in the Review of Particle Physics [4]. As we shall see, by employing various symmetries of QCD, it is possible to predict, at least qualitatively, many observable regularities of hadronic spectra and interactions.

The pre-QCD quark model of hadrons failed to explain why only mesons ($\bar{q}q$), baryons (qqq) and antibaryons ($\bar{q}\bar{q}\bar{q}$) are observed. Why, for example, are the diquark (qq) or four-quark ($qqqq$) bound states absent? In QCD, one immediately gets an explanation based on the colour confinement principle. The quark–antiquark and three-quark combinations can form colour-singlet states, whereas diquark and four-quark compounds are always colour-charged. The colour-singlet meson state is obtained by simply summing over the colour indices of the quark and antiquark. For example, the π^+ meson has the following flavour/colour structure:

$$|\pi^+\rangle = \frac{1}{\sqrt{3}} \sum_{i=1}^{3} |u^i \bar{d}_i\rangle \,. \tag{39}$$

The way baryons are built is less trivial. Two quarks are arranged in a coloured *diquark* state $\epsilon_{ijk}|q^j q^k\rangle$,

[3]The word 'hadron' was coined by Okun in Ref. [7] where he wrote: "It is reasonable to call strongly interacting particles hadrons, and the corresponding decays - hadronic. In Greek the word 'hadros' means 'large', 'massive', in contrast to the word 'leptos', which means 'small', 'light'. The term hadron refers to long-lived mesons and baryons, as well as to resonances".

where ϵ_{ijk} is totally antisymmetric with $\epsilon_{123} = 1$. This state has the colour SU(3)-transformation properties of an antiquark (with a colour index i). Hence, we can obtain a colour-neutral state combining a diquark with the third quark and summing over colours in the same way as in mesons. One important example of a three-quark baryon state[4] is Ω^- baryon with spin 3/2:

$$|\Omega\rangle = \frac{1}{\sqrt{6}} \sum_{i,j,k=1}^{3} \epsilon_{ijk} |s^i s^j s^k\rangle. \tag{40}$$

With respect to the flavour and spin this state is totally symmetric: all quark have the same flavour and parallel spins. Therefore, colour degrees of freedom provide the antisymmetry demanded by the Pauli principle for any bound state of fermions.

2.2 Quark model of hadrons

The hadron decompositions (39) and (40) resemble chemical formulae, displaying the content of a composite state. In reality hadrons are far more complicated than atoms and molecules, because the masses of the u, d, s quarks are smaller than the QCD long-distance scale: $m_{u,d} \ll \Lambda_{QCD}$ and $m_s < \Lambda_{QCD}$. Hence, light quarks are purely relativistic and the number of quarks, antiquarks and gluons within a hadron cannot be fixed. Since QCD is a quantum field theory, additional quark–antiquark pairs or gluons are created and annihilated inside the bound states *virtually*, i.e., within the time/distance intervals of $O(1/\Lambda_{QCD})$. As a result, the general decomposition of the physical pion state is not simply (39) but rather

$$|\,\pi^+\rangle = \;\; |\,u^i \bar{d}_i\rangle + \sum_{q=u,d,s,..} |\,u^i \bar{d}_i q^j \bar{q}_j\rangle + |\,u^k (\lambda^a)^i_k \bar{d}_i G^a\rangle + \ldots, \tag{41}$$

where G denotes a gluon, the Lorentz indices are not specified, and a summation over colour indices is implied. In the above sum, the first term represents a state with the minimal particle content (the so-called valence-quark content), and ellipses indicate all other multiparticle fluctuations. Naturally, all components of the pion state have to be colourless, with the same $u\bar{d}$ overall flavour, to obey the colour-neutrality and flavour conservation. We conclude that hadrons are, in general, many-body systems with relativistic constituents. Therefore, simple quantum-mechanical models with an interquark potential cannot fully describe pion or other light-quark hadrons.

Before attempting to solve the QCD dynamics, it is useful to apply the symmetries of QCD Lagrangian. The space–time (Lorentz–Poincare) invariance implies that the total angular momentum (or total spin) J of a hadron is a well defined and conserved quantum number. In addition P- and C- parities are conserved in QCD (as opposed to the electroweak theory). Therefore, for a given hadron, the spin-parity combination J^P (J^{PC} for flavour-neutral hadrons) is the next important signature after the mass. Spin parities are indicated for each observed hadron in its entry in [4].

To proceed in hadron spectroscopy, let us have a more detailed look at the mesons having the same flavour content $u\bar{d}$ as π^+. Each meson state is a complicated coherent decomposition similar to (41). Nevertheless, since J^P is conserved, it is sufficient to consider the valence-quark component to count all possible combinations of J^P starting from the lowest possible spin. The total angular momentum of the valence quark–antiquark state is a sum of the quark and antiquark spins plus the orbital angular momentum:

$$\vec{J} = \vec{S} + \vec{L}\ , \tag{42}$$

where $\vec{S} = \vec{s}_q + \vec{s}_{\bar{q}}$ is the total quark spin; $S = 1(0)$ if the individual spins are parallel (antiparallel). Accordingly, there are two possible states with $L = 0$ and $J = S$. One is with $J^P = 0^-$ (pion) and the other with $J^P = 1^-$ (ρ meson). The negative P parity attributed to these states is obtained from

[4]Note that if there were four colours of quarks, with the corresponding SU(4) symmetry, baryons would have been built from four quarks, with profound consequences for the physical world, e.g., atoms with fractional electric charge.

Table 2: Spectrum of the lowest $u\bar{d}$ states

	$S=0$		$S=1$	
L	J^P	meson	J^P	meson
0	0^-	$\pi(140)$	1^-	$\rho(770)$
1	1^+	$b_1(1235)$	0^+	$a_0(980)$
			1^+	$a_1(1260)$
			2^+	$a_2(1320)$

the following rule: $P = -(-1)^L$. Notice an additional minus which has to be added to account for the so-called 'intrinsic' P parity of the relativistic quark–antiquark system.

Turning to the states with $L = 1$, one encounters three mesons with $S = 1$ ($J^P = 0^+, 1^+, 2^+$) and one with $S = 0$ ($J^P = 1^+$). They are listed in Table 2 according to the classification of Ref. [4]. A similar counting can be done for $L = 2, 3, ...$, predicting $J > 2$ mesons. Some of them can be found in Ref. [4]. Generally, it is rather difficult to observe hadrons with higher spins. Having larger masses, these states have many decay channels and, therefore, a large total width, complicating their experimental identification in the form of a resonance. Baryons from u, d, s quarks with different J^P listed in Ref. [4] can also be interpreted, at least qualitatively, in terms of three-quark valence states with the orbital momentum L between diquark and the third quark.

The angular momentum is not the only source of excited hadron resonances. There are mesons which have the same J^P as π or ρ but a larger mass. These states are somewhat similar to the radially excited levels in a potential. For the pion a natural candidate of such excitation is the $\pi'(1300)$ state with $J^P = 0^-$, whereas the ρ meson has at least two experimentally established radially-excited partners with $J^P = 1^-$: $\rho'(1400)$ and $\rho''(1700)$ [4]. Note that because the orbital momentum L is not conserved in relativistic theory, the $L = 2$ state with $J^P = 1^-$ cannot be simply distinguished from the 'radial excitation' of the $L = 0$ state with the same spin-parity—another difficulty for the potential models of light-quark hadrons. Ultimately, one has to think in terms of purely relativistic extended objects, some kind of *quark–gluon strings* having a spectrum of 'radial' and J excitations. However, attempts to derive a string picture for hadrons directly from L_{QCD} have not been successful so far.

Quark–gluon interaction is flavour-independent. Therefore, given that a $u\bar{d}$ meson with a certain J^P exists, the mesons with the same J^P containing all possible quark–antiquark flavour combinations should also be observed. The flavour partners of the pion (ρ meson) with $J^P = 0^-$ ($J^P = 1^-$) are listed in Table 3. Almost all of them have been observed; the masses and other characteristics are given in Ref. [4]. The only temporary exceptions are the pseudoscalar $b\bar{b}$ meson (η_b) and the vector $\bar{b}c$ meson (B_c^*). These two states are not yet in Ref. [4], owing mainly to experimental reasons.

The *heavy quarkonia*, i.e., the mesons consisting of a heavy quark and antiquark ($\bar{c}c$ or $\bar{b}b$) are of special interest. Here the masses of interacting quarks are large enough compared to their characteristic energies within hadrons: $m_{b,c} \gg \Lambda_{QCD}$. In other words, heavy quarks are nonrelativistic objects with respect to QCD interactions. It is therefore possible to approximate the quark–gluon interactions with a nonrelativistic potential, putting the hadronic calculus on the safe ground of quantum mechanics. The Coulomb quark–antiquark potential $V(r) = \alpha_s/r$, derived from the one-gluon exchange, is valid at small distances. To provide quark confinement, a certain long-distance part of the potential, infinitely growing at $r \to \infty$ should also exist (e.g., oscillator or linear potential). This part of the potential cannot be directly calculated from L_{QCD} and is usually modelled and fitted to the observed quarkonium spectra. Importantly, numerical studies of QCD on the lattice confirm the existence of the confining potential force between heavy quark and antiquark.

Table 3: Pseudoscalar ($J^P = 0^-$) (upper lines) and vector ($J^P = 0^-$) (lower lines) mesons with different flavour content

	u	d	s	c	b
$\bar{u}$	π^0, η, η'	π^-	K^-	D^0	$\bar{B}^-$
	ρ^0, ω	ρ^-	K^{*-}	D^{*0}	$\bar{B}^{*-}$
$\bar{d}$	π^+	π^0, η, η'	$\bar{K}^0$	D^+	$\bar{B}^0$
	ρ^+	ρ^0, ω	$\bar{K}^{*0}$	D^{*+}	$\bar{B}^{*0}$
$\bar{s}$	K^+	K^0	η, η'	D_s	$\bar{B}_s$
	K^{*+}	$\bar{K}^{*0}$	ϕ	D_s^*	$\bar{B}_s^*$
$\bar{c}$	$\bar{D}^0$	D^-	$\bar{D}_s$	η_c	$\bar{B}_c$
	$\bar{D}^{*0}$	D^{*-}	$\bar{D}_s^*$	J/ψ	$\bar{B}_c^*$
$\bar{b}$	B^+	B^0	B_s	B_c	η_b
	B^{*+}	B^{*0}	B_s^*	B_c^*	Υ

2.3 Isospin

In addition to the exact colour and space–time symmetries, the QCD Lagrangian possesses approximate flavour symmetries originating from the pattern of quark masses. Since the latter are generated by some external mechanism, the flavour symmetries do not have fundamental roots in QCD. Nevertheless, they provide very important relations for hadron masses and hadronic amplitudes.

Let us start with the u and d quarks and rewrite the QCD Lagrangian, isolating these two flavours:

$$L_{QCD} = \bar{\psi}_u(iD_\mu\gamma^\mu - m_u)\psi_u + \bar{\psi}_d(iD_\mu\gamma^\mu - m_d)\psi_d + L_{glue} + L_{s,c,b,t}\,. \tag{43}$$

The smallness of the u and d masses, $m_u \sim m_d \ll \Lambda_{QCD}$, implies that their difference also is small:

$$m_d - m_u \ll \Lambda_{QCD}\,. \tag{44}$$

Neglecting this difference we use a new notation for the common u, d quark mass:

$$m_u \simeq m_d \simeq \tilde{m}\,. \tag{45}$$

In this approximation,

$$L_{QCD} \simeq L_{QCD}^{(u=d)} = \overline{\Psi}(D_\mu\gamma^\mu - \tilde{m})\Psi + L_{glue} + ..\,, \tag{46}$$

where a new, two-component spinor field (doublet) is introduced:

$$\Psi = \begin{pmatrix} \psi_u \\ \psi_d \end{pmatrix}\,, \quad \bar{\Psi} = (\bar{\psi}_u, \bar{\psi}_d)\,.$$

The theory described by the r.h.s. of (46) is not exactly QCD, but is very close to it. The new Lagrangian $L_{QCD}^{(u=d)}$ contains two degenerate quark flavours and has a symmetry with respect to the general phase rotations in the 'two-flavour space':

$$\Psi \to \Psi' = \exp\left(-i\sum_{a=1}^{3}\omega^a\frac{\sigma^a}{2}\right)\Psi, \quad \overline{\Psi} \to \overline{\Psi}' = \overline{\Psi}\exp\left(i\sum_{a=1}^{3}\omega^a\frac{\sigma^a}{2}\right)\,, \tag{47}$$

where ω^a are arbitrary (x-independent) parameters and $\sigma^1, \sigma^2, \sigma^3$ are the 2×2 Pauli matrices. The symmetry transformations (47) form a group SU(2)[5].

One has to emphasize that the approximate SU(2)-flavour symmetry emerges 'by chance', simply because the u and d quark masses turn out to be almost degenerate. Note also that e.m. interaction violates this symmetry, owing to different electric charges of u and d quarks. But this $O(\alpha_{em})$ effect is again small, at the same level of $\sim 1\%$, as the $O(\frac{m_d - m_u}{\Lambda_{QCD}})$ violation due to the quark mass difference.

The approximate degeneracy of u and d flavours manifests itself in hadrons. Replacing u quarks by d quarks or vice versa in a given hadron, yields a different hadron which has a very close mass and other properties. This qualitative prediction is nicely confirmed by the measured mass differences between proton (uud) and neutron (udd), $\pi^+(\bar{u}d)$ and $\pi^0([\bar{u}u - \bar{d}d]/\sqrt{2})$, $K^+(u\bar{s})$ and $K^0(d\bar{s})$, etc. The typical mass splittings for the $u \leftrightarrow d$ hadronic partners are at the level of a few MeV. Thus, QCD nicely explains the origin of *isospin* symmetry introduced by Heisenberg in the 1930s in nuclear physics to describe the similarities between the 'mirror' isotopes, obtained from each other by interchanging protons and neutrons. The second part in the word 'isospin' reflects the analogy with the electron spin symmetry, the degeneracy of the spin-up and spin-down electron states in quantum mechanics.

In fact, one introduces a similar formalism in QCD, attributing isospin $I = 1/2$ to the doublet of u and d quarks and treating these two flavours as 'up' and 'down' components with $I_3 = +1/2$ and $I_3 = -1/2$, respectively. The hadrons containing u and d quarks in different combinations form isomultiplets, with the isospin counting similar to the spin algebra in quantum mechanics. In the case of the proton and neutron, the diquark ud has isospin 0, therefore, adding u or d quark to the diquark, we get the nucleon isodoublet ($I = 1/2$) consisting of a proton with $I_3 = +1/2$ and a neutron with $I_3 = -1/2$. Another isodoublet is formed by K^+ and K^0, where the $\bar{s}$ quark, which has no isospin, is combined with u and d, respectively. In the same way, $D^0(u\bar{c})$ and $D^-(d\bar{c})$, or $B^+(u\bar{b})$ and $B^0(d\bar{b})$ build isodoublets.

Note that antiquarks have the opposite signs of I_3: $\bar{u}$ ($\bar{d}$) has $I_3 = -1/2$ ($I_3 = +1/2$). Combining the u and d quarks with their antiquarks, one gets four states. Three of them belong to isotriplet ($I = 1$):

$$u\bar{d}\,(I_3 = +1), \quad \frac{u\bar{u} - d\bar{d}}{\sqrt{2}}\,(I_3 = 0), \quad d\bar{u}\,(I_3 = -1)\,. \tag{48}$$

For example, pions (π^+, π^0 and π^-), as well as ρ mesons (ρ^+, ρ^0 and ρ^-) form isotriplets. The fourth quark–antiquark state is an isosinglet ($I = 0$):

$$\frac{u\bar{u} + d\bar{d}}{\sqrt{2}}\,, \tag{49}$$

which deserves a separate discussion. In general, the $u\bar{u}$ and $d\bar{d}$ states transform into each other via intermediate gluons. In mesons this transition takes place at long distances, owing to some nonperturbative mechanism, not necessarily described by diagrams with a fixed number of gluons. In any case, the transition amplitude has a characteristic scale of $O(\Lambda_{QCD})$, much larger than the mass difference $\simeq 2(m_u - m_d)$ between the $u\bar{u}$ and $d\bar{d}$ states. The $\bar{u}u$–$\bar{d}d$ degeneracy yields two orthogonal physical states, the $I_3 = 0$ component of the isotriplet (48) and the isosinglet (49). Turning to strange quarks, one encounters the second $I = 0$ state $\bar{s}s$. The same gluonic transition mechanism provides a mixing between (49) and the $\bar{s}s$ state. Now, the difference between the masses is not small, being of $O(2m_s - 2\tilde{m}) \sim \Lambda_{QCD}$. Hence, the amount of mixing depends on the magnitude of the $s\bar{s} \leftrightarrow (u\bar{u} + d\bar{d})$ transition amplitude. The latter is quite sensitive to the spin parity of the meson state. For example, the

[5]The number of independent parameters for SU(2) is determined in the same way as for SU(3) in Lecture 1: one counts the number of independent elements in the unitary 2×2 matrix with the unit determinant.

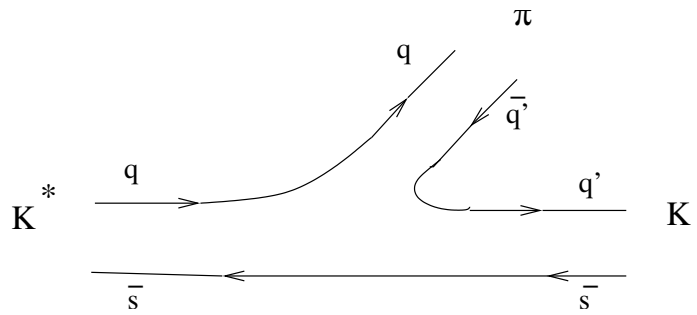


Fig. 10: Quark diagram for $K \to K\pi$ hadronic decays; the four possible combinations of $q = u, d$ and $q' = u, d$ correspond to the four decay modes related by isospin symmetry.

two isosinglet mesons with $J^P = 0^-$ shown in Table 3 have (up to small deviations) the following quark content:

$$\eta(547) \simeq \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2\bar{s}s)\,, \tag{50}$$

$$\eta'(958) \simeq \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + \bar{s}s)\,, \tag{51}$$

indicating that mixing in the O^- channel is large. For the $J^P = 1^-$ mesons the situation is completely different. From the quark content of the isosinglet mesons:

$$\omega(782) \simeq \frac{u\bar{u} + d\bar{d}}{\sqrt{2}}, \;\; \phi(1020) \simeq \bar{s}s\,, \tag{52}$$

one concludes that the transition from strange to nonstrange quark pairs in the 1^- state is suppressed.

Returning to the isospin symmetry, it is worth mentioning that it yields useful relations between hadronic amplitudes. To give just one simple example, let us consider the four observable hadronic decays of the K^* meson. The amplitudes of these decays are related via isospin symmetry, so that only one amplitude is independent:

$$A(K^{*0} \to K^+\pi^-) = -\frac{1}{\sqrt{2}}A(K^{*0} \to K^0\pi^0) = A(K^{*+} \to K^0\pi^+) = \frac{1}{\sqrt{2}}A(K^{*+} \to K^+\pi^0)\,. \tag{53}$$

To obtain these relations, one does not necessarily need to apply the formulae for the SU(2) group. In the isospin limit, all four decays are described by a single quark diagram, shown in Fig. 10, where the initial and final mesons are taken in the valence-quark state. Each individual decay mode has its own combination of u and d quarks to be substituted in this diagram. Furthermore, the coefficients $1/\sqrt{2}$ originate from the quark content of the π^0. Naturally, the hadronic amplitude attributed to the quark diagram in Fig. 10 is a nonperturbative, long-distance object and cannot be directly calculated in QCD. However, all we need are the amplitude relations between individual decay modes and not the value of the amplitude itself. Squaring the relations (53) and taking into account the phase space, one predicts the ratios of branching fractions confirmed by the experimental values given in Ref. [4].

2.4 SU(3) flavour symmetry

Another flavour symmetry, widely used in hadron phenomenology, is $SU(3)_{fl}$ corresponding to the limit of QCD with all three quarks, u,d and s, having equal masses. Since in reality m_s is only inessentially smaller than Λ_{QCD}, the magnitude of $SU(3)_{fl}$ violation in hadrons is not universal, depending on their quark content and quantum numbers.

Neglecting the mass differences between u,d,s quarks and introducing a common mass $\tilde{m}_3$ one approximates L_{QCD} as

$$L_{QCD} \simeq L_{QCD}^{(u=d=s)} = \overline{\Psi}_3 (D_\mu \gamma^\mu - \tilde{m}_3) \Psi_3 + L_{glue} + .. , \tag{54}$$

where Ψ_3 is a triplet:

$$\Psi_3 = \left(\begin{array}{c} \psi_u \\ \psi_d \\ \psi_s \end{array} \right) , \quad \bar{\Psi}_3 = (\bar{\psi}_u, \bar{\psi}_d, \bar{\psi}_s) .$$

The modified QCD with $L_{QCD}^{(u=d=s)}$ has a symmetry with respect to the transitions between three flavour states:

$$\Psi_3 \to \Psi_3' = \exp \left[-i \sum_{a=1}^{8} \omega^a \frac{\lambda^a}{2} \right] \Psi_3 , \quad \bar{\Psi}_3 \to \bar{\Psi}'_3 = \bar{\Psi}_3 \exp \left[i \sum_{a=1}^{8} \omega^a \frac{\lambda^a}{2} \right] . \tag{55}$$

Although physically, $SU(3)_{fl}$ and the fundamental $SU(3)$ colour have completely different origins, the group-theoretical formalism of both symmetries is the same. In particular, the eight λ^a matrices entering (55) are already given in (8).

The $SU(3)_{fl}$ symmetry is very helpful in 'organizing' the spectra of strange and nonstrange hadrons in multiplets. The light-quark meson multiplets are obtained by combining quark flavour-triplets and antiquark flavour-antitriplets. Without using the specific rules of $SU(3)$ algebra, which can be found in many textbooks, it is easy to figure out that the nine quark–antiquark states split into a singlet $\bar{\Psi}_3 \Psi_3$ and octet $\bar{\Psi}_3 \lambda^a \Psi_3$. The octet has its own isospin substructure[6]. The singlet-octet pattern provides a reasonable description for pseudoscalar mesons, in particular, the η' meson is close to the $SU(3)_{fl}$ singlet state (51). The isotriplet of pions (π^+, π^0, π^-), two isodoublets of kaons (K^+, K^0 and $\bar{K}^0, K^-$) and the isosinglet η, given by (50), together form an octet. However, this pattern is not universal. For example, in the case of vector mesons, ω and ϕ states in (52) are neither octets, nor singlets. To complete the counting of meson $SU(3)_{fl}$ multiplets, one has to mention also triplets and antitriplets of heavy–light mesons. For example, in the case of c quark, $\bar{D}^0, D^-, \bar{D}_s$ (D^0, D^-, D_s) form a triplet (antitriplet).

Generally, $SU(3)_{fl}$ works quite well for baryons, because their characteristic mass scale is a few times larger than Λ_{QCD}. Importantly, the spin-parity of the lowest baryon $SU(3)_{fl}$ multiplets is fixed, owing to the total antisymmetry of the baryon 'wave function' in the $SU(3)_{fl}$/spin/ colour coordinates required by Fermi statistics. There is an octet with $J = 1/2$ (including proton and neutron) and decuplet with $J = 3/2$. Let us, for instance, have a look at the latter. It contains the isoquadruplet ($I = 3/2$) of Δ resonances ($\Delta^{++}(uuu)$, $\Delta^+(uud)$, $\Delta^0(udd)$, $\Delta^-(ddd)$), the isotriplet of Σ resonances ($\Sigma^+(uus)$, $\Sigma^0(uds)$, $\Sigma^-(dds)$), the isodoublet of Ξ resonances ($\Xi^0(uss)$, $\Xi^-(dss)$) and the isosinglet $\Omega(sss)$. Consulting Ref. [4] for the masses of these baryons, one notices a distinct hierarchy: each constituent s quark adds an amount of $O(m_s)$ to the baryon mass.

Returning to the quark diagram in Fig. 10, we may now replace the s quark by a u or d quark. In the $SU(3)_{fl}$ limit the new diagram obtained after this replacement is equal to the one with the s quark, yielding relations between the $K^* \to K\pi$ and $\rho \to \pi\pi$ hadronic amplitudes, e.g.,

$$A(K^{*+} \to K^0 \pi^+) \simeq -\frac{1}{\sqrt{2}} A(\rho^+ \to \pi^0 \pi^+) . \tag{56}$$

Such relations are typically violated at the level of 20–30%, but are still useful from the phenomenological point of view.

[6] In mathematical terms $SU(2)_{isospin}$ is a subgroup of $SU(3)_{fl}$.

2.5 Heavy quark symmetry

To complete our survey of flavour symmeties, we now turn to the $\{c, b\}$ quark sector of the QCD Lagrangian. The fact that $m_{c,b} \gg \Lambda_{QCD}$, together with the flavour-independence of quark–gluon interactions, allows one to consider an interesting limit of L_{QCD} where both c and b quarks have infinitely heavy mass:

$$m_c \sim m_b \sim m_Q \to \infty \,. \tag{57}$$

At first sight, the limit is not justified, because in reality m_b is substantially larger than m_c. As we shall see, the fact that both masses are large turns out to be more important for QCD dynamics. Formally, in the limit (57) one can introduce a doublet of heavy-flavour fields

$$\Psi = \begin{pmatrix} \psi_c \\ \psi_b \end{pmatrix} \,, \quad \bar{\Psi} = (\bar{\psi_c}, \bar{\psi_b}) \,,$$

and rewrite the Lagrangian in a form invariant with respect to SU(2) rotations in the c, b flavour space:

$$L_{QCD} = \sum_{Q=c,b} \overline{\Psi}_Q (i D_\mu \gamma^\mu - m_Q) \Psi_Q + L_{glue} + L_{u,d,s} \,. \tag{58}$$

This particular form of the heavy-quark limit for L_{QCD} is, however, not convenient, because the heavy mass scale m_Q is still present explicitly. To understand why it is desirable to effectively remove that scale, let us consider the heavy-quark limit (57) for D or B meson. Since m_c and m_b have no direct relation to QCD, it makes sense to discuss a hypothetical heavy-light meson H with a mass m_H containing a heavy quark with an arbitrarily large mass m_Q. In the rest frame of H the constituent heavy quark stays almost at rest, providing a static source of colour charge which emits and absorbs gluon fields. The meson mass, to a good approximation is

$$m_H = m_Q + \bar{\Lambda} \,, \tag{59}$$

where Λ is the energy of the light quark–gluon cloud surrounding the heavy quark. The situation very much resembles the hydrogen atom where the total mass of the atom is a sum of an extremely large ($\sim$ GeV) proton mass and a small energy of the electron cloud ($\sim$ MeV). The essential point is that the electron itself is nonrelativistic. One can isolate the electron mass from the rest of the energy, introduce the Coulomb potential and kinetic energy, and eventually solve the equations of motion, determining the electron energy levels. In heavy hadrons the light-quark cloud is purely relativistic ($m_{u,d,s} < \Lambda_{QCD}$) and has a complicated long-distance nature.

Nevertheless, one essential feature is common for both bound states. In the atom the energy of the electron cloud does not depend on the proton mass. Likewise, in the heavy-light meson, $\bar{\Lambda}$ in (59) is approximately independent of m_Q. From atomic physics we know that the electron energy levels in hydrogen and deuterium coincide to a great precision. The fact that a deuteron is twice as massive as the proton does not play a role for the energy levels, because in both cases the atomic nuclei are static. What is important is that the electric charge does not change by switching from proton to deuteron. Similarly, in the H meson, $\bar{\Lambda}$ changes very little if one replaces m_Q by m_b or by m_c, because the colour charge of the heavy quark does not change. Thus, the heavy-flavour symmetry is in reality the symmetry between the light-quark remnants of the heavy hadrons, so that the heavy-quark mass scale indeed plays a secondary role.

To achieve a quantitative level, a special formalism of *heavy quark effective theory* (HQET) was developed for applications of QCD to heavy-light hadrons. One starts from L_{QCD} and introduces transformations which decouple the $\sim m_Q$ part of the heavy-quark field from the part which has the remnant momentum $\sim \Lambda_{QCD}$. Only the latter part strongly interacting with the light quark–gluon cloud is relevant for QCD dynamics. Therefore, in HQET one usually integrates out the heavy degrees of freedom and works with the Lagrangian containing a new effective quark field carrying the flavour of Q but no

mass. Not only the static limit (57) but also an expansion in powers of $1/m_Q$ corrections can be systematically treated in HQET. The field-theoretical aspects of heavy-mass expansion are nicely explained in the literature (see, for example, Refs. [8], [9]); I shall only focus on some important phenomenological consequences of heavy-flavour symmetry.

One famous example is the $B \to Dl\nu_l$ decay involving the weak $b \to c$ transition of the Standard Model. The unknown part of the decay amplitude is the hadronic matrix element

$$\langle D(p_D)|\bar{c}\gamma_\mu b|B(p_B\rangle \tag{60}$$

determined by the long-distance interactions involving the initial and final heavy quarks as well as the surrounding light quark–gluon 'cloud'. One chooses a special kinematical configuration, the 'zero recoil point' where the momentum transfer to the lepton pair is equal to

$$q^2 = (p_B - p_D)^2 = (m_B - m_D)^2 \, . \tag{61}$$

In the B meson rest system $p_B = (m_B, 0, 0, 0)$ this point corresponds to the final D meson at rest. In the heavy-quark limit the replacement of the b quark by the c quark does not change the hadronic state:

$$\langle D(p_D)| = \langle B(p_B)|$$

and the matrix element (61) reduces to a trivial normalization factor. One can therefore predict the decay amplitude in the zero recoil point up to $1/m_Q$ corrections. In fact there is a theorem stating that these corrections are even smaller and start from $O(1/m_Q^2)$, but to derive this and other important details one needs a full-scale HQET framework.

Without resorting to the effective theory, it is possible to understand the origin of another important symmetry emerging in the heavy-quark limit. In the hydrogen atom, the electron and proton have magnetic moments related to their spins and yielding interactions with the external magnetic fields or with each other (spin–spin interactions). The magnetic moments are inversely proportional to the masses, so that the proton magnetic moment plays no role for the electron energy levels. Each level is degenerate with respect to the proton spin direction. Since QED and QCD have very similar vector boson interactions with spin 1/2 particles, the spin $1/2$ quarks also have *chromomagnetic* moments and interact with the 'magnetic' parts of gluonic fields and with other quarks. For the heavy nonrelativistic quark the chromomagnetic moment is inversely proportional to the heavy quark mass m_Q. In the infinite mass limit the interaction vanishes, and hence the light-cloud energy $\bar{\Lambda}$ is independent of the spin orientation of the heavy quark.

One arrives at a new classification of heavy-light states based on this *heavy-quark spin symmetry*. Instead of adding together the orbital momentum and the total spin of quarks as we did in (42) it is more appropriate to introduce, for a $Q\bar{q}$ meson ($Q = c, b$, $q = u, d, s$), the total angular momentum of light degrees of freedom:

$$\vec{J}_{light} = \vec{L} + \vec{s}_q \, . \tag{62}$$

Adding the heavy-quark spin $s_Q = 1/2$ to J_{light} one gets degenerate doublets of heavy-light mesons with total angular momentum $J = J_{light} \pm 1/2$. At $L = 0$ one simply has $J_{light} = 1/2$ and therefore a doublet of mesons with $J^P = 1^-$ and $J^P = 0^-$ consisting of B and B^* (D and D^*) in the b quark (charm) sector. The mass differences within doublets are indeed very small [4]:

$$\delta_B = m_{B^*} - m_B = 47 \text{ MeV}, \;\; \delta_D = m_{D^*} - m_D = 142 \text{ MeV}, \tag{63}$$

indicating that the heavy-quark spin symmetry works quite well, especially for the heavier b quark. Taking into account that the mass differences are $\sim 1/m_Q$ effects, one expects that $\delta_B/\delta_D \simeq (m_c/m_b)$ which is also in accordance with (63) and (31). I leave as an exercise to show that at $L = 1$ there are two degenerate doublets: one with $J^P = 0^+, 1^+$ and another one with $J^P = 2^+, 1^+$.

2.6 Exotic hadrons

The colour confinement principle does not exclude hadronic states with an 'exotic' valence quark content, different from $q\bar{q}$ or qqq. Quarks, antiquarks, and gluons can be added together in any combination, e.g., $q\bar{q}G$, GG , $q\bar{q}q\bar{q}$ or $q\bar{q}qqq$, provided they are in a colour-neutral state. Since one cannot calculate the spectrum of hadrons in QCD with a good precision, predictions of exotic states are generally model-dependent. It is always problematic to distinguish an exotic hadron from the excitation of an ordinary hadron with the same J^P and flavour quantum numbers. Moreover, in this case one expects mixing between ordinary and exotic hadrons. For example, if there is a $J^P = 0^-$ state composed of two gluons GG (*glueball*), it should be mixed with η' to a certain degree, so that η' acquires a glueball component.

Therefore, the most interesting, 'smoking gun' signatures are the hadrons with exotic quantum numbers (flavour content and/or J^{PC}), forbidden for quark–antiquark mesons or three-quark baryons. For example it is impossible to arrange a flavour-neutral, quark–antiquark state with $J^{PC} = 1^{-+}$. The P and C (charge-conjugation) parities of a fermion–antifermion state are determined by the rules: $P = -(-1)^L$ and $C = (-1)^{L+S}$, so that $P = -1$ means $L = 0, 2, ...$ Hence, the only possibility to have $C = +1$ is $S = 0, 2, ...$ However, adding together even values of L and S one cannot get $J = 1$. On the other hand, adding one constituent gluon to a $q\bar{q}$ pair, one easily makes a 'hybrid' meson with $q\bar{q}G$ content and $J^{PC} = 1^{-+}$ quantum numbers. Searches for hybrid mesons are currently being carried out, but the experimental situation is not yet settled.

The recently observed narrow baryon resonance $\Theta(1540)$ decaying to K^+n Ref. [10] is another promising candidate for hadron exotics, a state $\bar{s}uudd$ with five valence constituents (*pentaquark*). Flavour symmetries are important model-independent tools to confirm/reject the experimental candidates for exotic resonances. In particular, an important task is to find the symmetry partners of these hadrons and to fill the relevant isospin and $SU(3)_{fl}$ multiplets (in the case of the pentaquark it is actually the $SU(3)_{fl}$ antidecuplet).

3 QCD AT SHORT DISTANCES

3.1 Probing short distances with electroweak quark currents

In this lecture we return to the quark–gluon interactions at large momentum transfers (short distances). In this region, practically at $Q \geq 1$ GeV, α_s is small, allowing one to apply the perturbative expansion. It is then possible to test QCD quantitatively, calculating various quark–gluon interaction processes at large Q and comparing the results with the available experimental data. Note, however, that the traditional way to study interactions by scattering one object on the other is not applicable to quarks and gluons. They simply are not available in free-particle states. One needs to trace quarks inside hadrons, where the long-distance forces are important. Take as an example the elastic pion–proton scattering at large momentum transfers (see Fig. 11). Here one has to combine the perturbative quark–quark scattering amplitudes (two-gluon exchange) with the 'wave functions' of quarks inside the initial and final hadrons. To obtain these functions one needs to go beyond perturbative QCD. Therefore, an unambiguous extraction of the perturbative amplitude from the data on the scattering cross section is not a realistic task.

The situation is not so hopeless, actually, since we have at our disposal electroweak bosons (γ, W, Z) interacting with quarks in a pointlike way. Electroweak interactions at large momentum transfer serve as external probes of short-distance dynamics. In Lecture 1, we already discussed one example: the quark–antiquark pair production in Z decay. To list all possible electroweak sources of quarks in a more systematic way, I start with the photon. Its interaction with the quark e.m. current was already given in (6), let me write it down again:

$$L_{em}(x) = -ej_\mu^{em} A^\mu \,, \tag{64}$$

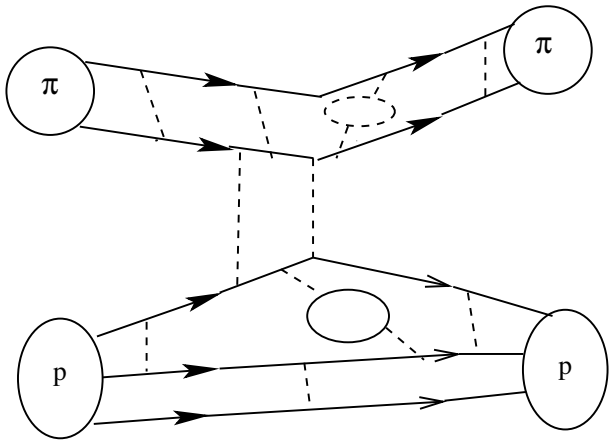


Fig. 11: A schematic view of pion–proton elastic scattering at large momentum transfers.

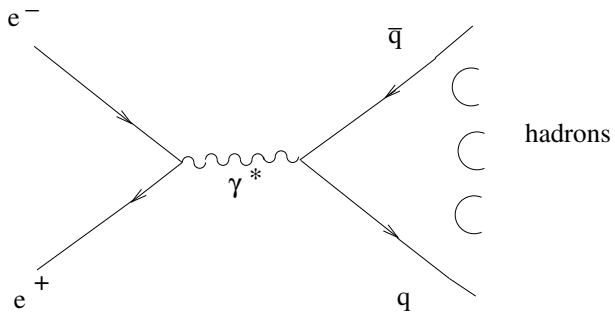


Fig. 12: The virtual-photon exchange diagram for $e^+e^- \to$ hadrons.

introducing a compact notation for the quark e.m. current:

$$j_\mu^{em} = \sum_{q=u,d,s,c,..} Q_q \bar{\psi}_q \gamma_\mu \psi_q \,, \tag{65}$$

where the summation over colour indices is not shown for the sake of brevity. The quark weak current entering the quark–W flavour-changing interaction:

$$L_W(x) = -\frac{g}{2\sqrt{2}} j_\mu^W W^\mu + c.c. \tag{66}$$

is more complicated and includes the CKM mixing matrix:

$$j_\mu^W(x) = (\bar{u}, \bar{c}, \bar{t})\gamma_\mu(1-\gamma_5) \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \tag{67}$$

Finally, the quark–Z interaction is

$$L^Z = -\frac{g}{2cos\theta_W} j_\mu^Z Z_\mu \,, \tag{68}$$

where the quark electroweak neutral current is a mixture of vector and axial-vector parts.

3.2 Perturbative QCD in $e^+e^- \to hadrons$

In e^+e^- annihilation at high energies the virtual photon provides a short-distance source of quark–antiquark pairs. This process is hadron-free in the initial state. The photon-exchange diagram is depicted

in Fig. 12 (for simplicity I ignore the additional Z-exchange diagram). The experimentally measured total cross-section $\sigma_{tot}(e^+e^- \to h)$ depends on one kinematical variable $s = (p_{e^-} + p_{e^+})^2$. The virtual timelike photon transfers its energy $\sqrt{s}$ to the hadronic state. At very large $\sqrt{s} \gg \Lambda_{QCD}$ and $\sqrt{s} \gg m_q$, the initial pair of quarks is produced at an average distance of $O(1/\sqrt{s})$, much smaller than the typical hadronic distance scale $1/\Lambda_{QCD}$. Owing to asymptotic freedom of QCD, gluonic interactions of the produced quark pair are suppressed by small α_s. Hence, e^+e^- annihilation at high energies provides an almost pointlike source of quasi-free quark pairs. At long distances the created quarks and antiquarks are inevitably converted into some hadronic state. Since in the total cross section the summation is done over all hadronic states produced at a given energy, the total probability of hadronization sums up to a unit. Hence, at $\sqrt{s} \to \infty$ the hadronic cross section is well approximated by the cross section of the free quark–antiquark pair production:

$$\sigma_{tot}^{(e^+e^- \to h)}(s) \simeq \sum_{q=u,d,s,..} \sigma^{(e^+e^- \to q\bar{q})}(s)\,, \tag{69}$$

summed over all quark flavours with $m_q \ll \sqrt{s}$. This, so-called *parton model* approximation for $e^+e^- \to h$ is confirmed by experimental data. Moreover, the majority of events saturating the cross section at high $\sqrt{s}$ consists of two distinct hadronic jets originating from the initial quark pair.

The way we obtained (69) may seem too qualitative and a bit 'hand-waving'. In the following, we shall derive the asymptotic cross section (69) in a more rigorous way. In this derivation several important concepts will be introduced, to be used in discussing further topics covered by these lectures.

We start with the formal definition of the total cross section:

$$\sigma_{tot}^{(e^+e^- \to h)}(s) = \frac{1}{2s} \sum_{h_n} \left| \langle h_n \mid T \mid e^+e^- \rangle \right|^2 , \tag{70}$$

where the sum over the final hadronic states h_n includes phase-space integration and implies summation over spins (polarizations). The matrix elements $\langle f|T|i\rangle \equiv T_{fi}$ (in (70) $|i\rangle = |e^+e^-\rangle$ and $\langle f| = \langle h_n|$) determine the general S-matrix of the theory:

$$S_{fi} \equiv \langle f|S|i\rangle = \delta_{fi} + iT_{fi}\,. \tag{71}$$

The usual representation of the S-matrix in terms of Lagrangian has the time-ordered exponential form:

$$S = T\left\{ \exp\left[i \int d^4x \left(L_{QCD}(x) + L_{QED}(x) \right) \right] \right\} , \tag{72}$$

where L_{QED} includes e.m. interactions of quarks and leptons. Furthermore, the unitarity of the S matrix is used:

$$SS^\dagger = 1\,, \tag{73}$$

or

$$\sum_n \langle f|S|n\rangle\langle n|S^\dagger|i\rangle = \delta_{fi}\,. \tag{74}$$

From now on we consider the forward scattering $f = i$. Replacing $\langle n|S^\dagger|i\rangle = \langle i|S|n\rangle^*$ and substituting (71) in (74), one obtains the unitarity relation for T_{ii} (the optical theorem):

$$2\ \mathrm{Im}T_{ii} = \sum_n |T_{in}|^2\,. \tag{75}$$

To apply this universal relation to e^+e^- scattering, we take $|i\rangle = |e^+e^-\rangle$ with four-momentum $q = p_{e^+} + p_{e^-}$, so that $q^2 = s > 0$. Furthermore, we choose $|n\rangle = |h_n\rangle$ restricting the set of intermediate states by hadronic states. As a result we obtain a rigorous relation between the amplitude of the forward

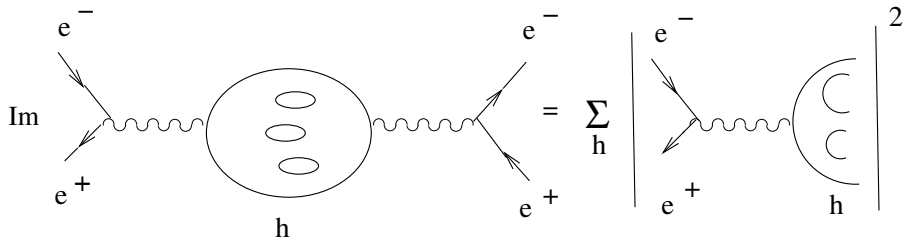


Fig. 13: Unitarity relation for $e^+e^- \to$ hadrons.

$e^+e^- \to h \to e^+e^-$ scattering via hadronic intermediate states, $T_{ii} = \mathcal{A}^{(e^+e^- \to h \to e^+e^-)}$, and the sum over the squared $e^+e^- \to h_n$ amplitudes. The latter sum, according to (70) is proportional to the total hadronic cross section. The optical theorem (75) takes the form:

$$2\,\mathrm{Im}\,\mathcal{A}^{(e^+e^- \to h \to e^+e^-)}(s) = \sum_{h_n} |\langle h_n|T|e^+e^-\rangle|^2 = 2s\sigma_{tot}^{(e^+e^- \to h)}(s)\,. \tag{76}$$

Diagrammatically, this relation is represented in Fig. 13. The amplitude

$$\mathcal{A}^{(e^+e^- \to h \to e^+e^-)}(q^2) = \frac{e^4}{(q^2)^2}(\bar\psi_e\gamma^\mu\psi_e)(\bar\psi_e\gamma^\nu\psi_e)\Pi_{\mu\nu}(q) \tag{77}$$

contains the photon propagators and the products of electron and positron spinors in both initial and final states, written according to the QED Feynman rules. The nontrivial part of this amplitude, denoted as $\Pi_{\mu\nu}$, describes the $j_\mu^{em} \to h \to j_\nu^{em}$ transition, and is called the *correlation function* (or correlator) of quark currents. The formal expression for this object reads:

$$\Pi_{\mu\nu}(q) = i\int d^4x\; e^{iqx}\langle 0|T\{j_\mu^{em}(x)j_\nu^{em}(x)\}|0\rangle\,. \tag{78}$$

Owing to the conservation of e.m. current ($\partial^\mu j_\mu^{em} = 0$), the correlation function depends on one invariant amplitude:

$$\Pi_{\mu\nu}(q) = (-g_{\mu\nu}q^2 + q_\mu q_\nu)\Pi(q^2)\,. \tag{79}$$

Substituting (79) in (77) and taking the imaginary part from both sides, we obtain

$$\mathrm{Im}\,\mathcal{A}^{(e^+e^- \to h \to e^+e^-)}(s) = -\frac{e^4}{s}(\bar\psi_e\gamma^\mu\psi_e)(\bar\psi_e\gamma_\mu\psi_e)\mathrm{Im}\,\Pi(s)\,. \tag{80}$$

It is convenient to normalize the hadronic cross section to the $e^+e^- \to \mu^+\mu^-$ cross section known from QED. One can literally repeat the derivation done above, taking instead of hadronic states the $\mu^+\mu^-$ states: $|n\rangle = |\mu^+\mu^-\rangle$. The resulting relations are quite similar to (76), (77) and (80):

$$2\,\mathrm{Im}\,\mathcal{A}^{(e^+e^- \to \mu^+\mu^- \to e^+e^-)}(s) = \sum_{\mu^+\mu^-} |\langle \mu^+\mu^-|T|e^+e^-\rangle|^2 = 2s\sigma^{(e^+e^- \to \mu^+\mu^-)}(s)\,, \tag{81}$$

$$\mathcal{A}^{(e^+e^- \to \mu^+\mu^- \to e^+e^-)}(q^2) = \frac{e^4}{(q^2)^2}(\bar\psi_e\gamma^\rho\psi_e)(\bar\psi_e\gamma^\lambda\psi_e)\Pi_{\rho\lambda}^{(\mu)}(q)\,, \tag{82}$$

and

$$\mathrm{Im}\,\mathcal{A}^{(e^+e^- \to \mu^+\mu^- \to e^+e^-)}(s) = -\frac{e^4}{s}(\bar\psi_e\gamma^\mu\psi_e)(\bar\psi_e\gamma_\mu\psi_e)\mathrm{Im}\,\Pi^{(\mu)}(s)\,, \tag{83}$$

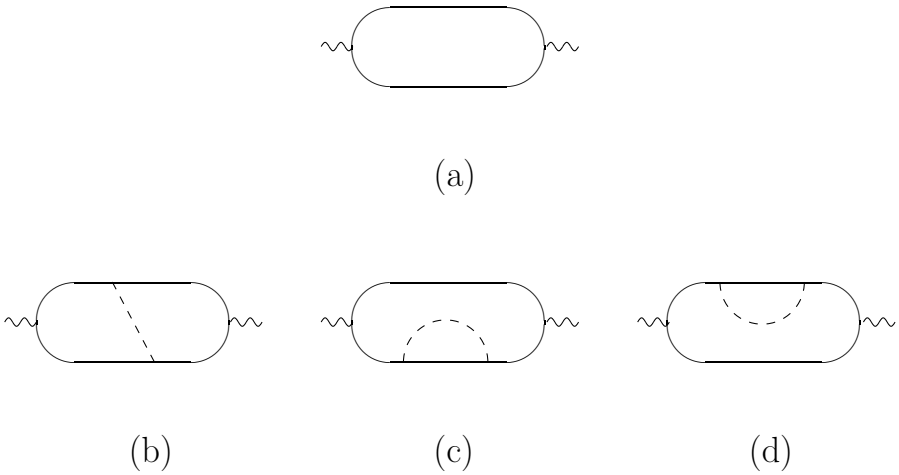


Fig. 14: Diagrams corresponding to the perturbative contributions to the correlation function of two quark currents: (a) the leading-order loop and (b)–(d) the $O(\alpha_s)$ corrections. Wavy lines denote external currents with 4-momentum q, solid lines quarks, and dashed lines gluons.

where the muonic correlation function $\Pi^{(\mu)}$ is nothing but a simple 2-point muon-loop diagram. Furthermore, the cross section in (81) taken from QED textbooks, reads:

$$\sigma^{(e^+e^-\to\mu^+\mu^-)}(s) = \frac{4\pi\alpha_{em}^2}{3s}\,. \tag{84}$$

Dividing the hadronic unitarity relation (76) by the muonic one (81) and using (80) and (83), we obtain a useful ratio:

$$\frac{\mathrm{Im}\,\Pi(s)}{\mathrm{Im}\,\Pi^{(\mu)}(s)} = \frac{\sigma_{tot}^{(e^+e^-\to h)}(s)}{\sigma^{(e^+e^-\to\mu^+\mu^-)}(s)} \equiv R(s)\,. \tag{85}$$

The next key point in our derivation is the analysis of the correlation function $\Pi(q^2)$ at spacelike $q^2 < 0$. At large $|q^2| = Q^2 \gg \Lambda_{QCD}^2$, the long-distance domain in the space–time integral in (78) is suppressed by the strongly oscillating exponent, and the short distances/times $|\vec{x}| \sim x_0 \sim 1/Q$ dominate. This justifies using QCD perturbation theory with $\alpha_s(Q) \to 0$ at $Q \to \infty$. The leading-order asymptotically free result is given by the 2-point quark-loop diagram [Fig. 14(a)], and the next-to-leading corrections are determined by $O(\alpha_s)$ two-loop diagrams [Fig. 14(b), (c), (d)]. Calculation of these diagrams yields (at $m_q \ll Q$):

$$\Pi(q^2) = -\frac{1}{4\pi^2}\left(\sum_q Q_q^2\right)\log\left(\frac{-q^2}{\mu^2}\right)\left(1+\frac{\alpha_s}{\pi}\right) + O\left(\frac{1}{\epsilon}\right) + \mathrm{const}\,, \tag{86}$$

where each flavour q contributes with the same expression and a coefficient Q_q^2. The natural scale for α_s is q^2. In recent years, thanks to tremendous calculational efforts, the $O(\alpha_s^2)$ and even $O(\alpha_s^3)$ corrections to Π have been calculated; also the loop diagrams for the massive quark are known with a high accuracy.

In the final stage of our derivation the calculated function $\Pi(q^2)$ at $q^2 < 0$ is related to $\mathrm{Im}\Pi(s)$ at positive s. One employs Cauchy's theorem for the function $\Pi(z)$ obtained from $\Pi(q^2)$ by analytically continuing the real variable q^2 to the complex values, $q^2 \to z$:

$$\Pi(q^2) = \frac{1}{2\pi i}\int\limits_C dz \frac{\Pi(z)}{z-q^2}\,. \tag{87}$$

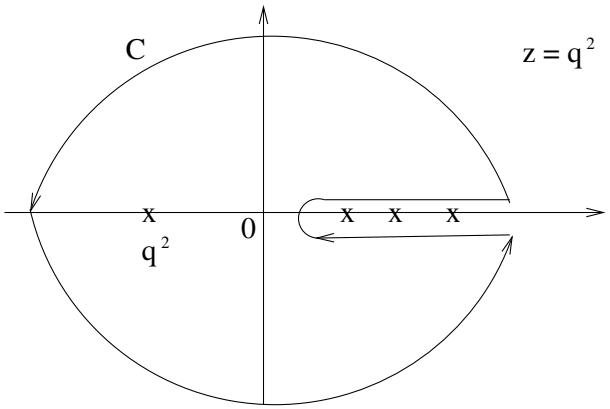


Fig. 15: The integration contour in (87). The crosses on the positive real axis indicate singularities of $\Pi(q^2)$.

The integration contour C is shown in Fig. 15. It circumvents the singularities of the function $\Pi(z)$, i.e., the points/regions where Im $\Pi \neq 0$. According to (76) and (80) the location of singularities at real $q^2 > 0$ is determined by the masses of resonances and/or the thresholds of multiparticle hadronic states produced in $e^+e^- \to h$; the lowest one is at $s_{min} = 4m_\pi^2$ corresponding to the threshold of the lightest two-pion state. Subdividing the contour C into 1) a large circle with the radius R, 2) an infinitely small semicircle $\tilde{C}$ surrounding s_{min}, and 3) two straight lines from s_{min} to R, we can rewrite the integral in terms of three separate contributions:

$$\Pi(q^2) = \frac{1}{2\pi i}\int\limits_{|z|=R} dz \frac{\Pi(z)}{z-q^2} + \frac{1}{2\pi i}\int\limits_{s_{min}}^{R} dz \frac{\Pi(z+i\delta)-\Pi(z-i\delta)}{z-q^2} + \frac{1}{2\pi}\int\limits_{\tilde{C}} dz \frac{\Pi(z)}{z-q^2}\,. \tag{88}$$

Suppose the function decreases at $|q^2| \to \infty$, $\Pi(q^2) \sim 1/|q^2|^\lambda$, where $\lambda > 0$. Then the first integral vanishes at $R \to \infty$. Taking an infinitely small semicircle, one makes the third integral also vanishing. Furthermore, since there are no singularities of $\Pi(z)$ at Re $z < s_{min}$, the integrand in the second integral reduces to the imaginary part: $\Pi(q^2 + i\delta) - \Pi(q^2 - i\delta) = 2i$ Im $\Pi(q^2)$ (due to Schwartz reflection principle). Finally, we obtain the desired *dispersion relation*

$$\Pi(q^2) = \frac{1}{\pi}\int\limits_{s_{min}}^{\infty} ds \frac{\text{Im}\,\Pi(s)}{s-q^2-i\delta} \tag{89}$$

with the l.h.s. calculated in QCD in a form of perturbative expansion (86) and the r.h.s. related to the cross section via (85):

$$\text{Im}\,\Pi(s) = R(s)\text{Im}\,\Pi^{(\mu)}(s)\,. \tag{90}$$

The quantity $R(s)$ is directly measurable in e^+e^- experiments. It remains to determine Im $\Pi^{(\mu)}(s)$. Knowing the answer for the quark loop diagram in Fig. 14a it is very easy to write down the expression for the muon loop. Since we neglect masses in both diagrams, the only difference is the factor 3, from summing up the colour states in the quark loop. This factor is naturally absent for the muon loop. From (86), taking imaginary part, one obtains

$$\text{Im}\,\Pi^{(\mu)}(s) = \frac{1}{12\pi}\,. \tag{91}$$

Finally, to guarantee the convergence of the dispersion integral (89),[7] let us differentiate both parts in q^2:

$$\frac{d\Pi(q^2)}{dq^2} = \frac{1}{\pi}\int\limits_{s_{min}}^{\infty} ds \frac{\mathrm{Im}\,\Pi(s)}{(s-q^2)^2}\,. \tag{92}$$

where (86) gives for the l.h.s.

$$\frac{d\Pi(q^2)}{dq^2} = \left(\sum_q^{n_f} Q_q^2\right)\left(-\frac{1}{4\pi^2 q^2}\right)\,. \tag{93}$$

Note that divergent and constant terms disappeared after differentiation and play no role in our derivation. The final form of the dispersion relation is

$$\frac{3\left(\sum\limits_q^{n_f} Q_q^2\right)}{-q^2}\left(1+\frac{\alpha_s}{\pi}+...\right) = \int\limits_{s_{min}}^{\infty} ds \frac{R(s)}{(s-q^2)^2}\ , \tag{94}$$

where ellipses denote higher-order in α_s corrections. The fact that (94) is valid at $(-q^2)\to\infty$ unambiguously fixes the constant limit of $R(s)$ at large s:

$$R(s) \to 3\sum_q^{n_f} Q_q^2\,, \tag{95}$$

where n_f indicates that R includes all 'active' quark flavours, for which the condition $\sqrt{s} \gg m_q$ is fulfilled. Finally, we notice that (95) coincides with the parton model prediction (69), taking into account that the free-quark and muon-pair cross sections differ only by the colour factor times the quark charge squared:

$$\sigma^{(e^+e^-\to q\bar{q})}(s) = 3Q_q^2\sigma^{(e^+e^-\to\mu^+\mu^-)}(s)\,.$$

Importantly, QCD not only reproduces the parton model prediction for $R(s)$ but also provides perturbative corrections, as well as predicting the integral (94) over $R(s)$.

The experimental data collected in various regions of $\sqrt{s}$ nicely confirm (95). According to Fig. 16 taken from Ref. [4], the ratio $R(s)$ approaches first the constant value $R \simeq 2$ at energies $\sim$ 2–3 GeV, above the region of vector meson resonances ρ,ω,ϕ (and below the charm threshold). That is exactly the value anticipated from (95) for $n_f = 3$. Well above charmonium resonances, a new constant level is achieved: $R = 2 + 3Q_c^2 = 10/3$. And finally, $R = 10/3 + 3Q_b^2 = 11/3$ is settled at energies above Υ resonances, where all five quark flavours are in their asymptotic regime. Actually, the current data on $R(s)$ are so precise that one should also include small α_s corrections to R_{QCD}.

There are other similar inclusive observables calculable in QCD, among them the total widths $\Gamma_{tot}(Z\to hadrons)$ and $\Gamma_{tot}(W\to hadrons)$. They have the same status as $R(s)$, but a fixed scale m_Z or m_W instead of $\sqrt{s}$. One has also to mention an interesting and well developed sub-field of perturbative QCD related to the jet and/or heavy-quark production in e^+e^- and hadron collisions at high energies. The underlying short-distance quark–gluon processes are successfully traced in the experimentally observed multijet structure of the final state. Naturally, hadrons cannot be completely avoided, because, after all, quarks and gluons hadronize. In fact, hadronization in jet physics is nowadays a somewhat routine procedure described by QCD-oriented models (e.g., the Lund model integrated within PYTHIA [11]). At lower scales, $Q \sim$ 1–2 GeV, inclusive decays of τ-lepton are among useful tools to study perturbative QCD (see, for example, Ref. [12]).

[7] For brevity, I avoid a longer derivation which includes some special mathematical construction (subtractions).

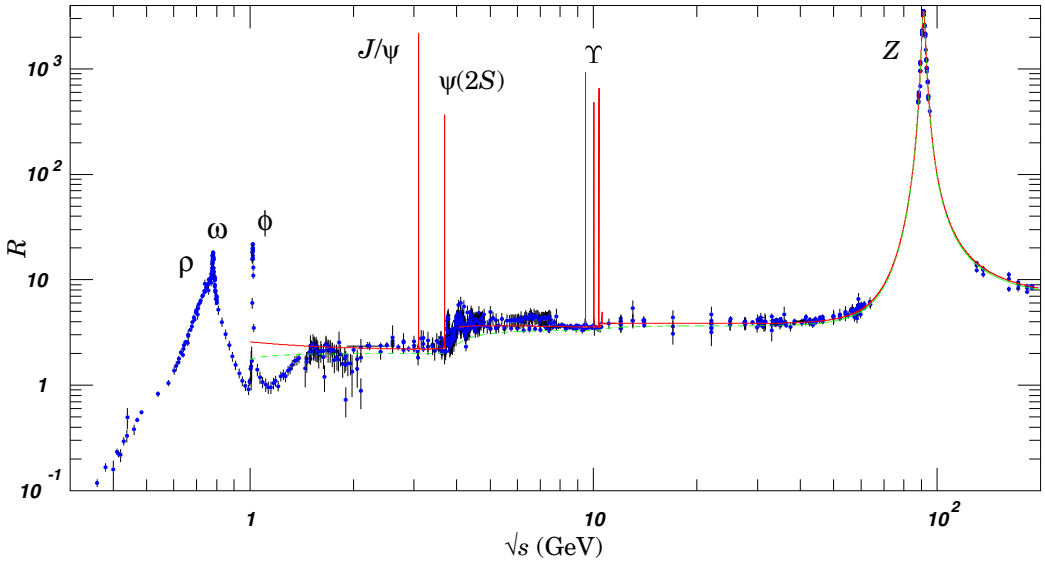


Fig. 16: Ratio R from Ref. [4].

3.3 Deep inelastic scattering and operator-product expansion

The processes of lepton–nucleon *deep inelastic scattering* (DIS), $lN \to lh$ and $\nu_l N \to lh$ ($l = e, \mu$), are among the best testing grounds of QCD. The long-distance structure of the initial nucleon makes these processes more complicated than $e^+e^- \to hadrons$. For definiteness, let us consider the electron–nucleon scattering mediated by the virtual photon γ^* with the 4-momentum q, whereas p is the nucleon 4-momentum. To measure the DIS cross section, one only has to detect the final electron. In the nucleon rest frame the invariant variables $Q^2 = -q^2$ and $\nu = q \cdot p$ are related to the initial and final energies and the scattering angle of the electron: $Q^2 = 4EE' sin^2\theta$, $\nu = (E - E')m_N$.

The specific kinematical region $Q^2 \sim \nu \gg \Lambda^2_{QCD}$ has to be chosen to reveal the spectacular effect of asymptotic freedom. In this region the experimentally measured DIS cross section, normalized to the cross section of the electron scattering on a free pointlike quark σ_{point}, depends essentially on the ratio $x = Q^2/2\nu$. This effect (*Bjorken scaling*), was first interpreted in terms of a beautiful phenomenological model suggested by Feynman. One considers the reference frame with large nucleon momentum p, so that $-\vec{q}$ gives the longitudinal direction. Neglecting long-distance binding forces between quarks, the initial nucleon is represented as a bunch of free constituents (*partons*): quarks, antiquarks, and gluons moving in the longitudinal direction[8]. For simplicity the quark masses and transverse momenta as well as the nucleon mass are neglected, in comparison with Q^2 and ν. The electron scatters on one of the nucleon constituents (excluding gluons, of course, because they are electrically neutral) which has the momentum fraction ξ, so that after the pointlike collision the quark 4-momentum is $p\xi + q$. The massless quark has to remain on-shell, $(\xi p + q)^2 = 0$, therefore $2\xi(p \cdot q) + q^2 = 0$ and

$$\xi = Q^2/2\nu = x\,. \tag{96}$$

The cross section is then represented as a sum of all possible 'elementary' processes integrated over ξ

$$d\sigma^{(eN\to eh)}(Q^2,\nu) = \int_0^1 \sum_{i=u,\bar{u},d,\bar{d}...} f_i(\xi) d\sigma^i_{point}(Q^2,\nu)\delta(\xi - x)d\xi = \sum_i d\sigma^i_{point}(Q^2,\nu) f_i(x)\,, \tag{97}$$

[8] We refer here to a generic picture of the nucleon, where all possible multiparticle components are coherently added to the valence three-quark state, similar to (41) for the pion.

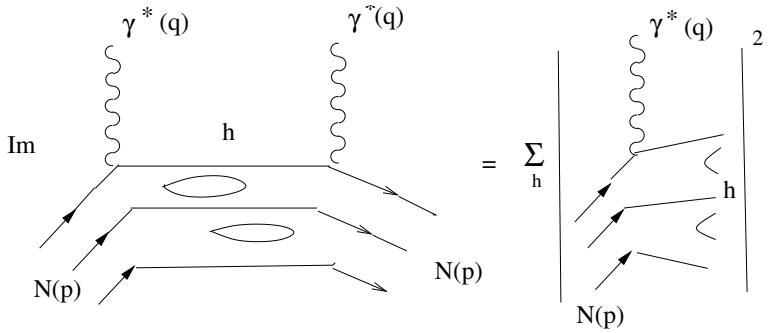


Fig. 17: Unitarity relation for DIS cross-section.

where the δ-function takes into account (96). The sum goes over all quark and antiquark species inside the proton, and $f_i(\xi)$ are the *parton distributions* defined as the probabilities to find the i-th constituent with the momentum fraction ξ.

In QCD, the scaling behaviour (97) corresponds to the asymptotic-freedom ($\alpha_s = 0$) approximation. Switching on the perturbative quark–gluon interactions one predicts logarithmic corrections $\sim \log(Q^2)$ to this formula which are nicely reproduced by experiment. There are many excellent reviews and lectures where DIS in perturbative QCD are discussed (see, for example, Ref. [13]). Here we shall concentrate on one essential aspect: separation of long- and short-distance effects.

Let us approach DIS from the quantum-field theory side, as we did above for $e^+e^- \to hadrons$. Omitting for simplicity the initial and final electrons, one can represent the hadronic part of the DIS cross section in a form of $\gamma^* N \to h$ cross section. Employing the unitarity relation (75) with $|i\rangle = |\gamma^* N\rangle$, one is able to relate the DIS cross section to the $\gamma^*(q)N(p) \to \gamma^*(q)N(p)$ forward-scattering amplitude (see Fig. 17):

$$2\ \mathrm{Im}\ \mathcal{A}^{(\gamma^* N \to h \to \gamma^* N)}(q,p) = \sum_{h_n} |\langle h_n|T|\gamma^* N\rangle|^2 \sim d\sigma^{(\gamma^* N \to h)}(Q^2,\nu)\,. \tag{98}$$

The amplitude

$$\mathcal{A}^{(\gamma^* N \to h \to \gamma^* N)}(q,p) = \epsilon^\mu(q)\epsilon^\nu(q) T_{\mu\nu}(q,p) \tag{99}$$

contains the photon polarization vectors multiplied by a new purely hadronic object

$$T_{\mu\nu}(p,q) = i\int d^4x e^{iqx}\langle N(p)|T\{j^{em}_\mu(x) j^{em}_\nu(0)\}|N(p)\rangle\,, \tag{100}$$

which resembles the correlation function of the two currents we introduced above, but, instead of vacuum, has nucleons in the initial and final states:

One can prove that at large Q^2 and ν the dominant contribution to the above integral stems from small space–time intervals $x^2 \sim 1/Q^2 \sim 1/\nu$ [9]. In other words, the points of absorption and emission of the virtual photon are located close to the light-cone $x^2 = 0$. The process takes place in the asymptotic freedom regime, that is a single quark absorbs the photon and penetrates quasi-freely at small x^2 before emitting the photon. All other contributions, for example, with different quarks emitting and absorbing initial and final photons are suppressed by powers of $1/Q^2, 1/\nu$. To describe the free-quark propagation from x to 0 we use the quark propagator

$$\langle 0|\psi_q(x)\bar\psi_q(0)|0\rangle = \int d^4p \left(\frac{p_\alpha\gamma^\alpha}{p^2}\right) e^{-ipx} = \frac{i x_\alpha \gamma^\alpha}{2\pi^2 (x^2)^2}\,, \tag{101}$$

[9] Note that the light-cone dominance $x^2 = x_0^2 - \vec{x}^2 \sim 0$ is a more general condition than the small-distance/time dominance $x_0 \sim \vec{x} \sim 0$ which takes place in $e^+e^- \to hadrons$.

neglecting the quark mass. Substituting the e.m. currents in (99) in terms of quark fields, contracting the fields of the propagating quark and using (101) we obtain

$$T_{\mu\nu}(p,q) = i\int d^4x e^{iqx} \sum_{q=u,d,s,..} \langle N(p)|T\{\bar\psi_q(x)\gamma_\mu\psi_q(x)\bar\psi_q(0)\gamma_\nu\psi_q(0)\}|N(p)\rangle$$
$$= i\int d^4x e^{iqx}\frac{ix_\alpha}{2\pi^2(x^2)^2}\sum_{q=u,d,s,..}\langle N(p)|\bar\psi_q(x)\gamma_\mu\gamma_\alpha\gamma_\nu\bar\psi_q(0)|N(p)\rangle + \dots \quad . \tag{102}$$

In this expression, where only the leading term is shown, the calculable short-distance part (the quark propagator) is separated from the long-distance part which is represented by the quark–antiquark matrix element taken between nucleon states. This hadronic matrix element is a complicated object which has to be either determined from experiment or calculated using methods beyond perturbative QCD which will be discussed in the next two lectures. I skip the derivation of the cross section formula from the imaginary part of $T_{\mu\nu}$ which allows one to relate the matrix element introduced in (102) with the parton distributions. Also α_s corrections can be systematically calculated; they contain important and observable $\log Q^2$ effects. Important is that the long-distance matrix element (or parton distributions) are universal characteristics of the nucleon and they do not change if the short-distance part of the process changes (e.g., by switching to neutrino–nucleon scattering $W^*N \to h_c$ where the charmed quark is produced). To summarize, in DIS it is possible to separate the short-distance domain by choosing the appropriate kinematical region and defining convenient physical observables. The short-distance quark–gluon interactions are calculable in a form of perturbation theory in α_s, whereas the long-distance part is parametrized in terms of universal hadronic matrix elements. The procedure of approximating the product of current in (102) by a quark–antiquark operator and separating short and long distances is called *operator-product expansion* (OPE) and is implicitly or explicitly used in almost any perturbative QCD treatment of hadronic processes.

4 LONG-DISTANCE DYNAMICS AND QCD VACUUM

4.1 Vacuum condensates

QCD at short distances does not essentially help in understanding the long-distance dynamics of quarks and gluons. From the short-distance side we only know that the running coupling $\alpha_s(Q)$ increases at low momentum scales and eventually diverges at $Q \sim \Lambda_{QCD}$ (see Fig. 7). Is the growth of α_s the only dynamical reason for confinement? It is not possible to answer this question remaining within the perturbative QCD framework, because the language of Feynman diagrams with propagators and vertices is already not applicable at $\alpha_s \sim 1$. QCD in the nonperturbative regime is currently being studied using other methods, first of all, lattice simulations. From these studies, there is a growing confidence that long-distance dynamics is closely related to the nontrivial properties of the physical vacuum in QCD.

For a given dynamical system, vacuum is a state with the minimal possible energy. Evidently, in QCD the vacuum state contains no hadrons, because creating any hadron always costs a certain amount of energy. In that sense, the QCD vacuum has to be identified with the $\langle 0 \mid$ state in the correlation function (78). Given that the vacuum state contains no hadrons does not yet mean that it is completely empty. There could be quantum fluctuations of quark and gluon fields with nonvanishing densities. The existence of vacuum fields is manifested, e.g., by *instantons*, special solutions of QCD equations of motion having a form of localized dense gluonic fields (for an introductionary review on instantons see, for example, Ref. [14]). Lattice QCD provides an independent evidence for quark/gluon fields in the vacuum. Without going into further theoretical details, I shall rather concentrate on the phenomenology of vacuum fields in QCD. We shall see that properties of hadrons are influenced by the existence of quark and gluon vacuum fluctuations with nonvanishing average densities, the so-called *vacuum condensates*.

Formally, in the presence of vacuum fields, the matrix elements of quark and gluon field operators between the initial $|0\rangle$ and final $\langle 0|$ states are different from zero. The combinations of fields have to

obey Lorentz-invariance, colour gauge symmetry, and flavour conservation, so that the simplest allowed composite operators are

$$O_3 = \bar{\psi}_{q\,i}\psi_q^i, \;\; O_4 = G^a_{\mu\nu}G^{a\mu\nu}, \;\; O_5 = \bar{\psi}_{q\,i}(\lambda^a)^i_k\sigma_{\mu\nu}G^{a\mu\nu}\psi_q^k,$$
$$O_6 = \left[\bar{\psi}_{q\,i}(\Gamma^a)^i_k\psi_q^k\right]\left[\bar{\psi}_{q'\,j}(\Gamma^a)^j_l\psi_{q'}^l\right], \tag{103}$$

where $\sigma_{\mu\nu} = (i/2)[\gamma_\mu, \gamma_\nu]$ and Γ^a are various combinations of Lorentz and colour matrices. The indices at O_d reflect their dimension d in GeV units. Furthermore, if the operators are taken at different 4-points, care should be taken of the local gauge invariance. For instance, the quark–antiquark nonlocal matrix element has the following form:

$$\langle 0 \mid \bar{\psi}_q(x)[x,0]\psi_q(0) \mid 0\rangle\,, \tag{104}$$

where $[x,0] = \exp\left[ig_s\int_0^1 dv x^\mu A^a_\mu(vx)(\lambda^a/2)\right]$ is the so-called gauge factor. Only the matrix elements with the light quarks $q = u, d, s$ are relevant for the nonperturbative long-distance dynamics. A pair of heavy c (b) quarks can be created in vacuum only at short distances/times of $O(1/2m_c)$ ($O(1/2m_b)$), that is, perturbatively.

Without fully solving QCD, very little could be said about vacuum fields, in particular about their fluctuations at long distances which have typical 'wavelengths' of $O(1/\Lambda_{QCD})$. Thus, we are not able to calculate the matrix element (104) explicitly, as a function of x. It is still possible to investigate the vacuum phenomena in QCD applying certain approximations. One possibility is to study the average local densities. The vacuum average of the product of quark and antiquark fields,

$$\langle 0 \mid \bar{\psi}_{q\,k}\psi_q^k \mid 0\rangle \equiv \langle\bar{q}q\rangle\,, \tag{105}$$

corresponds to the $x \to 0$ limit of the matrix element (104). The simplest vacuum density of gluon fields is

$$\langle 0 \mid G^a_{\mu\nu}G^{a\mu\nu} \mid 0\rangle \equiv \langle GG\rangle\,. \tag{106}$$

Because of translational invariance, both $\langle\bar{q}q\rangle$ and $\langle GG\rangle$ are independent of the 4-coordinate. These universal parameters are usually called the densities of quark and gluon condensates, respectively. As we shall see in the following subsection, the nonvanishing quark condensate drastically influences the symmetry properties of QCD.

4.2 Chiral symmetry and its violation in QCD

Let us return to the isospin symmetry limit (46) of the QCD Lagrangian:

$$L^{(u=d)}_{QCD} = \overline{\Psi}(iD_\mu\gamma^\mu - \tilde{m})\Psi + L_{glue} + \ldots\;. \tag{107}$$

Since $\tilde{m} \simeq m_u \simeq m_d \ll \Lambda_{QCD}$, a reasonable approximation is to put $\tilde{m} \to 0$, so that u- and d-quark components of the Ψ-doublet become massless.

Each Dirac spinor can be decomposed into the left- and right-handed components,

$$\psi_q = \frac{1+\gamma_5}{2}\psi_q + \frac{1-\gamma_5}{2}\psi_q \equiv \psi_{q\,R} + \psi_{q\,L}\,, \tag{108}$$

where, by definition, the left-handed (right-handed) quark has an antiparallel (parallel) spin projection on its 3-momentum. Similarly, for the conjugated fields one has:

$$\bar{\psi}_q = \bar{\psi}_q\frac{1-\gamma_5}{2} + \bar{\psi}_q\frac{1+\gamma_5}{2} \equiv \bar{\psi}_{q\,R} + \psi_{q\,L}\,. \tag{109}$$

Rewriting Ψ in terms of the left- and right-handed components, we obtain the following decomposition of the Lagrangian (107) in the massless limit:

$$L^{(u=d)}_{QCD} = \overline{\Psi}_R\, iD_\mu\gamma^\mu\Psi_R + \overline{\Psi}_L\, iD_\mu\gamma^\mu\Psi_L + L_{glue} + \ldots\;. \tag{110}$$

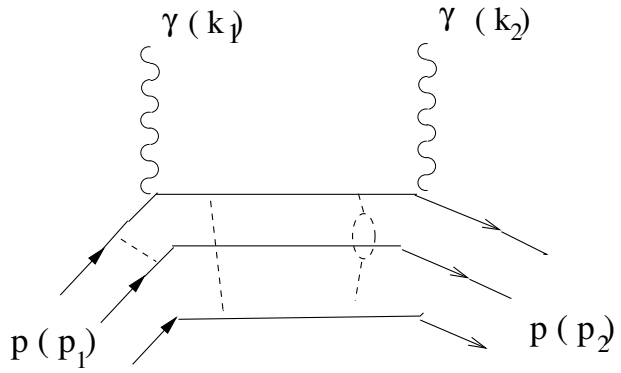


Fig. 18: One of the diagrams describing photon–proton scattering. The second diagram is obtained by interchanging the photon lines.

The quark–gluon interaction term in $L_{QCD}^{(u=d)}$ is now split into two parts, $g_s\overline{\Psi}_R\gamma_\mu A^{\mu a}(\lambda^a/2)\Psi_R$ and $g_s\overline{\Psi}_L\gamma_\mu A^{\mu a}(\lambda^a/2)\Psi_L$, so that quarks conserve their *chirality* (left- or right-handedness) after emitting/absorbing an arbitrary number of gluons. Importantly, also the interaction of quarks with photons has the same property (take the e.m. interaction (64) and decompose it into L and R parts in the same way as above). In the massless (*chiral*) limit, quarks of left and right chiralities propagate and interact independently from each other. In fact, it is possible to introduce two independent isospin $SU(2)$ transformations, separately for L and R fields:

$$\Psi_L \to \Psi'_L = \exp\left(-i\frac{\sigma^a}{2}\omega_L^a\right)\Psi_L\,, \qquad \Psi_R \to \Psi'_R = \exp\left(-i\frac{\sigma^a}{2}\omega_L^a\right)\Psi_R\,. \tag{111}$$

Restoring the mass in $L_{QCD}^{(u=d)}$ leads to the violation of chiral symmetry. The Lagrangian mass term can be represented as an effective transition between left- and right-handed quarks:

$$\tilde{m}\overline{\Psi}\Psi = \tilde{m}\left(\overline{\Psi}_L\Psi_R + \overline{\Psi}_R\Psi_L\right)\,. \tag{112}$$

Having in mind the smallness of the u,d-quark masses, one naturally expects that an approximate chiral symmetry manifests itself in the observable properties of hadrons. In reality, the symmetry is violated quite substantially, as the two following examples demonstrate.

The first one is a real hadronic process: the photon scattering on a longitudinally-polarized (e.g. left-handed) proton. One of the diagrams is shown in Fig. 18. This process is a very complicated mixture of quark–photon and quark–gluon interactions at long distances, determined by the quark structure of the proton. Importantly, all these interactions obey an approximate chiral symmetry at the Lagrangian level. Hence, the amplitude of the proton-chirality flip is expected to be very small, $O(m_{u,d}/\Lambda_{QCD}) \sim 1\%$ of the total scattering amplitude. To check this conjecture, let me consider the case when the initial photon energy is much smaller than the proton mass. In this case, it is possible to approximate the proton with a point-like particle. The $\gamma p \to \gamma p$ amplitude is then simply obtained from the $\gamma e \to \gamma e$ amplitude in QED (Compton effect), replacing the electron spinors and propagators by the proton ones:

$$A(\gamma p \to \gamma p) \simeq e^2\bar{u}(p_2)\left[\gamma_\alpha\frac{(p_1+k_1)_\mu\gamma^\mu + m_p}{(p_1+k_1)^2-m_p^2}\gamma_\beta + \gamma_\beta\frac{(p_1-k_2)_\mu\gamma^\mu + m_p}{(p_1-k_2)^2-m_p^2}\gamma_\beta\right]u_L(p_1)\epsilon_2^\alpha\epsilon_1^\beta, \tag{113}$$

where $\epsilon_{1,2}$ are the polarization vectors of the initial and final photons with the 4-momenta k_1 and k_2, respectively. Without even completing the calculation, we notice that the $\sim m_p$ terms in the amplitude flip the initial proton chirality $L \to R$, whereas the $\sim \gamma_\mu$ terms preserve chirality. Importantly, the contributions of both types are of the same order, determined by the scale m_p, indicating that chiral symmetry for the photon–proton scattering is broken at the 100% level.

To present the second example of the chiral symmetry violation, I start from the correlation function

$$\Pi_{\mu\nu} = i \int d^4x e^{iqx} \langle 0|T\{j_\mu(x) j_\nu(0)\}|0\rangle = (-g_{\mu\nu} q^2 + q_\mu q_\nu)\Pi(q^2)\,, \tag{114}$$

very similar to the one introduced in Lecture 3, but containing a slightly different quark current:

$$j_\mu = \frac{1}{\sqrt{2}}(\bar\psi_u \gamma_\mu \psi_u - \bar\psi_d \gamma_\mu \psi_d)\,, \tag{115}$$

which produces the $I = 1$ and $J^P = 1^-$ quark–antiquark states. Note that this current is conserved, $\partial_\mu j^\mu = 0$, even if $m_{u,d} \neq 0$. We then follow the same derivation as in Lecture 3 and obtain the dispersion relation (89) for $\Pi(q^2)$. The only change is in the imaginary part (90), where now only the states with $I = 1$ contribute to the total cross section (ρ meson and its radial excitations, the two-pion state and other states with an even number of pions), so that the ratio $R(s)$ has to be replaced by

$$R^{(I=1)}(s) = \frac{\sigma_{tot}(e^+e^- \to h(I=1))}{\sigma(e^+e^- \to \mu^+\mu^-)}\,. \tag{116}$$

In parallel, we consider the correlation function

$$\Pi^5_{\mu\nu} = i \int d^4x e^{iqx} \langle 0|T\{j_{\mu 5}(x) j_{\nu 5}(0)\}|0\rangle \tag{117}$$

of two axial-vector currents with $I = 1$:

$$j_{\mu 5} = \frac{1}{\sqrt{2}}(\bar\psi_u \gamma_\mu \gamma_5 \psi_u - \bar\psi_d \gamma_\mu \gamma_5 \psi_d)\,. \tag{118}$$

This current is conserved only in the chiral symmetry ($m_{u,d} = 0$) limit:

$$\partial^\mu j_{\mu 5} = \frac{1}{\sqrt{2}}(2 m_u \bar\psi_u \gamma_5 \psi_u - 2 m_d \bar\psi_d \gamma_5 \psi_d)\,. \tag{119}$$

Decomposing the correlation function (117) in two tensor structures:

$$\Pi^5_{\mu\nu}(q) = -g_{\mu\nu}\Pi^5_1(q^2) + q_\mu q_\nu \Pi_5(q^2)\,, \tag{120}$$

we notice that in the chiral limit there is only one independent invariant amplitude:

$$\Pi_5(q^2) = \frac{\Pi^5_1(q^2)}{q^2}\,. \tag{121}$$

The dispersion relation for Π_5 has the same form as (89):

$$\Pi_5(q^2) = \frac{1}{\pi} \int\limits_{s^5_{min}}^{\infty} ds \frac{\text{Im}\,\Pi_5(s)}{s - q^2 - i\delta}\,, \tag{122}$$

where s^5_{min} is the corresponding threshold. In order to determine the imaginary part via a relation similar to (90), we introduce a slightly artificial cross section $\sigma_{tot}(e^+e^- \to Z^A \to h(I=1))$ of hadron production mediated by a Z boson coupled to the axial-vector current. Then,

$$\text{Im}\,\Pi_5(s) = \frac{1}{12\pi} R^{(I=1)}_5(s)\,, \tag{123}$$

where

$$R^{(I=1)}_5(s) = \frac{\sigma^{(e^+e^- \to Z^A \to h(I=1))}_{tot}(s)}{\sigma^{(e^+e^- \to Z^A \to \mu^+\mu^-)}(s)}\,. \tag{124}$$

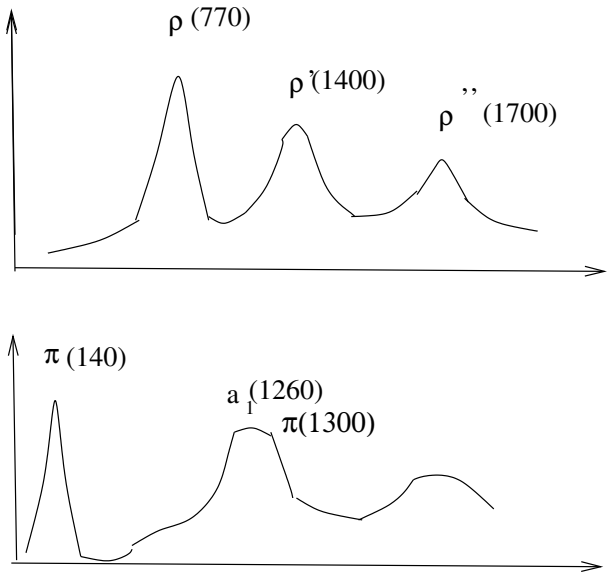


Fig. 19: A schematic pattern of the resonances produced by the vector (above) and axial-vector (below) quark currents with $I = 1$.

Both $\Pi(q^2)$ and $\Pi_5(q^2)$ can be calculated at large $|q^2|$ from the same 2-point quark-loop diagrams shown in Fig. 14. In the chiral limit, the only difference is in two extra γ_5 matrices present in $\Pi^5_{\mu\nu}$. Hence, in perturbative QCD

$$\Pi_5(q^2) = \Pi(q^2)\,, \tag{125}$$

at any order in α_s. This equation is trivial for the leading-order loop diagrams [Fig. 14(a)]. Using Dirac algebra, it is easy to check that the extra γ_5 matrices cancel each other ($\gamma_5^2 = 1$) in the absence of masses in the propagators. Furthermore, each gluon line inserted in the loop brings two more γ-matrices which do not influence that cancellation.

From (125) follows the equation of two dispersion relations (89) and (122). To keep the dispersion integrals convergent, we differentiate them n times at some $q^2 < 0$. The result is:

$$\int\limits_{s_{min}}^{\infty} ds \frac{R_5(s)}{(s-q^2)^n} = \int\limits_{s_{min}}^{\infty} ds \frac{R(s)}{(s-q^2)^n}\,. \tag{126}$$

Note that even in the presence of quark masses the corrections to (125) and to (126) are very small, $O(m^2_{u,d})$. If q^2 is not very large, $|q^2| \sim 1$ GeV^2, the integrals in (126) are dominated by the contributions of the low-mass hadronic states to the corresponding $e^+e^- \to h$ cross-sections. In the case of a vector current, the states are ρ meson and its radial excitations. Hence, $R(s)$ represents a resonance curve with the peaks located at $s = m^2_\rho$, $m^2_{\rho'}$ etc. (see Fig. 19). The validity of (126) at arbitrary (but large) $q^2 < 0$ implies that also $R_5(s) \simeq R(s)$. We then expect the spectrum of resonances generated by the axial-vector current to resemble, in gross features, the ρ spectrum. However, experimental data [4] reveal a completely different picture. The lowest resonances in the axial-vector channel are the pion with $J^P = 0^-$, the axial meson $a_1(1260)$ and the radial excitation $\pi(1300)$. Summarizing, there is clear evidence, based on the observed properties of hadrons, that chiral symmetry in QCD is violated much stronger than expected from L_{QCD}.

An additional source of the chiral symmetry violation is provided by the quark condensate. Decomposing the quark and antiquark fields in (105) in the left-handed and right-handed components,

$$\langle 0|\bar\psi_{q\,i}\psi^i_q|0\rangle = \langle 0|(\bar\psi_{qR} + \bar\psi_{qL})(\psi_{qR} + \psi_{qL})|0\rangle = \langle 0|\bar\psi_{qR}\psi_{qL} + \bar\psi_{qL}\psi_{qR})|0\rangle \neq 0\,, \tag{127}$$

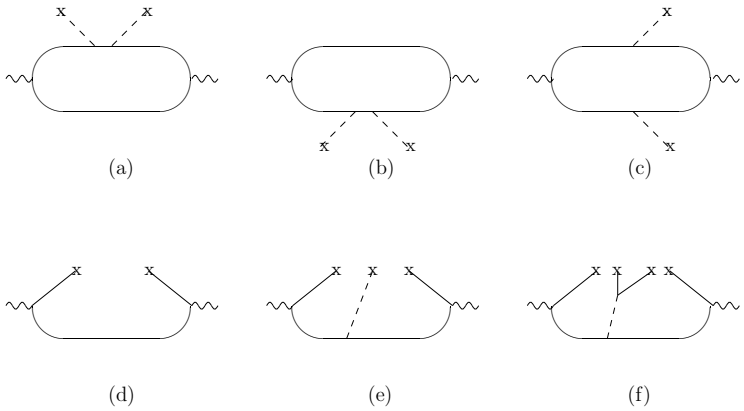


Fig. 20: Diagrams used to calculate condensate contributions to the correlation functions of two quark currents. Lines with crosses denote vacuum fields.

we realize that the condensate causes vacuum transitions between quarks of different chiralities. Hence, in QCD one encounters a *spontaneous broken* chiral symmetry, a specific situation when the interaction (in this case L_{QCD}) obeys the symmetry (up to the small $O(m_{u,d})$ corrections), whereas the lowest-energy state (QCD vacuum) violates it[10]. We conclude that in order to correctly reproduce the properties of hadrons and hadronic amplitudes (e.g., correlation functions), one has to take into account the vacuum fields, in particular, the quark condensate.

4.3 Condensate contributions to correlation functions

Let us return to the correlation functions (114) and (117), restoring nonzero u and d quark masses, that is, working with full QCD. The vacuum quark–gluon fields generate new contributions to Π or Π_5. In addition to the perturbative loop diagrams in Fig. 14, there are diagrams shown in Fig. 20, where gluons, quarks, and antiquarks penetrate to long distances, being absorbed and emitted by vacuum fluctuations. The vacuum fields have characteristic momenta of $O(\Lambda_{QCD})$. Therefore, if the momentum scale in the correlator is large, $Q = \sqrt{-q^2} \gg \Lambda_{QCD}$, one can approximate the vacuum state by a set of constant fields. In other words, the virtual quark propagating at short distances/times between the points x and 0, cannot 'resolve' the long-distance fluctuations of vacuum fields and percepts them in a form of averaged static fields. This approximation makes the calculation of diagrams in Fig. 20 straightforward. In addition to Feynman rules of perturbative QCD for virtual quarks and gluons, one has to form all possible combinations (103) of vacuum fields and replace them by the corresponding condensate densities, so that the 4-momenta of the crossed lines on these diagrams are neglected. For example, the product of quark and antiquark vacuum fields in Fig. 20(d) has to be replaced by $\langle\bar{q}q\rangle$. More details on these calculations can be found, for example, in the review [15]. The result for the vector-current correlator has the following schematic form:

$$\Pi(q^2) = \Pi^{pert}(q^2) + \sum_{d=3,4,\ldots} C_d(q^2)\langle 0|O_d|0\rangle\,, \tag{128}$$

where the first term on the r.h.s. corresponds to the perturbative diagrams in Fig. 14, whereas the sum contains the contributions of vacuum condensates with dimensions d obtained from the diagrams in Fig. 20. To compensate the growing dimension of the operators O_d, the coefficients C_d contain increasing powers of $1/Q$. Thus, the condensate contributions die away at $Q \to \infty$ and do not alter the perturbative

[10]That is quite similar to the electroweak sector of the Standard Model [1], where the electroweak gauge symmetry is spontaneously broken by the nonvanishing vacuum average of the Higgs field.

asymptotics of the correlator given by $\Pi^{pert}(q^2)$. Furthermore, as soon as we work at relatively large $Q \gg \Lambda_{QCD}$, it is possible to retain only a few first terms in the sum, that is, neglect diagrams with more than 3–4 vacuum fields emitted from the virtual quarks in the correlation function.

For the axial-vector correlation function $\Pi_5(q^2)$ one obtains an expression similar to (128) with the same perturbative part (up to very small corrections of $O(m^2_{u,d})$), but with different coefficients C_d at certain condensate terms. The most important deviation from the vector-current case is in the value and sign of C_6, i.e., in the 4-quark condensate terms. Thus, the addition of condensate effects leads to an explicit violation of (125). It is then not surprising that hadron resonances contributing to $R(s)$ and $R_5(s)$ in (126) are different.

Furthermore, the method of correlation functions allows one to reproduce an important relation for the pion mass, explaining the smallness of m_π. Note that from the point of view of the naive quark model, π and ρ mesons differ only by orientations of quark spins. Why is then $m_\pi \ll m_\rho$ and, moreover, $m_\pi < \Lambda_{QCD}$? We consider the correlation function similar to (117), but for simplicity, containing charged axial currents:

$$\Pi^5_{\mu\nu} = i \int d^4x e^{iqx} \langle 0|T\{j_{\mu 5}(x) j^\dagger_{\nu 5}(0)\}|0\rangle \; = -g_{\mu\nu}\Pi'_5(q^2) + q_\mu q_\nu \Pi_5(q^2)\,, \tag{129}$$

where $j_{\mu 5} = \bar{u}\gamma_\mu\gamma_5 d$. This current is the part of the Standard Model weak current (67), responsible for the $u \to d$ transition, e.g., the $\pi \to \mu\nu_\mu$ decay. The hadronic matrix element which determines this decay,

$$\langle 0|\bar{u}\gamma_\mu\gamma_5 d|\pi^+(q)\rangle = iq_\mu f_\pi\,, \tag{130}$$

is parametrized via the pion decay constant f_π which plays an essential role in our analysis. It is convenient to multiply (129) by $q_\mu q_\nu/q^2$, forming a combination of invariant amplitudes:

$$-\frac{q^\mu q^\nu}{q^2}\Pi^5_{\mu\nu} = \Pi'_5(q^2) - q^2\Pi_5(q^2) \equiv \tilde{\Pi}_5(q^2)\,.$$

Note that $\tilde{\Pi}_5 = 0$ at $m_{u,d} = 0$ and, in particular, the perturbative part of $\tilde{\Pi}_5$ vanishes as $O(m^2_{u,d})$. Concerning the nonperturbative part, the only first-order in the $m_{u,d}$ contribution is given by the quark condensate diagram in Fig. 20(d)

$$\tilde{\Pi}_5(q^2) = -\frac{(m_u + m_d)(\langle\bar{u}u\rangle + \langle\bar{d}d\rangle)}{q^2} + O(m^2_{u,d})\,. \tag{131}$$

To proceed, we use for $\tilde{\Pi}_5$ the dispersion relation of the type (122). To obtain the imaginary part, one has to return to the unitarity relation (75), identifying $|i\rangle$ with $|0\rangle$ and T_{ii} with the correlation function. The result is:

$$2\ \mathrm{Im}\,\tilde{\Pi}_5(s) = -\frac{q^\mu q^\nu}{q^2}\sum_{h_n}\langle 0|j_{\mu 5}|h_n\rangle\langle h_n|j^\dagger_{\nu 5}(q)|0\rangle\,. \tag{132}$$

Importantly, only pseudoscalar states contribute to the above sum, because the matrix elements for the axial-vector mesons a_1 and its excitations vanish, being proportional to the transverse polarization vectors of these mesons:

$$q^\mu\langle 0|j_{\mu 5}|a_1\rangle \sim q^\mu\epsilon^{a_1}_\mu = 0\,.$$

Using (132) and the definition (130) we obtain the following expression for the dispersion relation:

$$\tilde{\Pi}_5(q^2) \;=\; \frac{1}{2\pi}\int\limits_{m_\pi^2}^{\infty}\frac{ds}{s - q^2 - i\delta}\left\{-\frac{q^\mu q^\nu}{q^2}(f_\pi q_\mu)(f_\pi q_\nu)\int\frac{d^3p_\pi(2\pi)}{2E_\pi}\delta^{(4)}(p_\pi - q)\right\}_{q^2=s} + ...$$

$$= \int\limits_{m_\pi^2}^{\infty} \frac{ds}{s-q^2-i\delta} \left\{ -f_\pi^2 s\delta(s-m_\pi^2) \right\} + \ldots \ , \tag{133}$$

where the one-pion state contribution is shown explicitly (with $p_\pi^2 = m_\pi^2$) including the phase space proportional to δ-function. Ellipses denote excited pions and multiparticle states with the same quantum numbers. Integrating out the δ function and substituting (131) in the l.h.s. we obtain

$$-\frac{(m_u+m_d)(\langle\bar{u}u\rangle + \langle\bar{d}d\rangle)}{q^2} + O(m_{u,d}^2) = -\frac{f_\pi^2 m_\pi^2}{m_\pi^2 - q^2} - \sum_{\pi'} \frac{f_{\pi'}^2 m_{\pi'}^2}{m_{\pi'}^2 - q^2} \, , \tag{134}$$

where the sum over the higher state is also shown schematically. To fulfil this equation at large q^2 one has to demand that $m_\pi^2 \sim m_u + m_d$. Simultaneously, the decay constants of excited states have to obey: $f_{\pi'} \sim m_u + m_d$, otherwise the $1/q^2$ asymptotics of both parts in (134) is violated at $O(m_{u,d})$. We then reproduce the well known Gell-Mann–Oakes–Renner relation

$$-(m_u+m_d)(\langle\bar{u}u\rangle + \langle\bar{d}d\rangle) + O(m_{u,d}^2) = f_\pi^2 m_\pi^2 \ . \tag{135}$$

This relation reflects the special nature of the pion in QCD. The anomalously small pion mass is not accidental and is closely related to the spontaneous chiral-symmetry breaking via condensate. If a symmetry in quantum field theory is broken spontaneously, there should be massless states (Nambu–Goldstone particles), one per each degree of freedom of broken symmetry. The three pions, π^+, π^-, π^0 play a role of massless Nambu–Goldstone particles in QCD. In other words, due to the specific structure of QCD vacuum fields, the amount of energy needed to produce a pion state tends to zero. The fact that pions still have small nonvanishing masses is due to the explicit violation of chiral symmetry via u, d quark masses.

How large is the quark condensate density? Using $m_{\pi^\pm} \simeq 140$ MeV, $f_\pi = 131$ MeV [4] and taking the u, d quark mass values from (31), renormalizing them at $Q = 1$ GeV, one obtains from (135), typically,

$$\langle\bar{q}q\rangle(\mu = 1 \text{ GeV}) \simeq (-240 \pm 10 \text{ MeV})^3 \, , \tag{136}$$

where we assume isospin symmetry for the condensates $\langle\bar{q}q\rangle \simeq \langle\bar{u}u\rangle \simeq \langle\bar{d}d\rangle$. Not surprisingly, the estimated value is in the ballpark of Λ_{QCD}! Being not a measurable physical quantity, the condensate density is a scale-dependent parameter. Since the r.h.s. of (135) is determined by the hadronic parameters f_π, m_π, which are both scale-independent, the running of the quark condensate should compensate the running of the quark mass given in (37), that is:

$$\langle\bar{q}q\rangle(Q) = \langle\bar{q}q\rangle(Q_0) \left(\frac{\alpha_s(Q)}{\alpha_s(Q_0)} \right)^{-\gamma_0/\beta_0} \ . \tag{137}$$

4.4 Gluon condensate

The gluon condensate density is another important characteristic of nonperturbative QCD. This parameter cannot be easily estimated from correlation functions with light quarks, because the latter are dominated by quark condensates. A very useful object, sensitive to the gluon condensate, is the correlation function of c-quark currents:

$$\Pi_{\mu\nu}^c = i \int d^4x e^{iqx} \langle 0|T\{j_\mu^c(x) j_\nu^c(0)\}|0\rangle = (-g_{\mu\nu}q^2 + q_\mu q_\nu)\Pi^c(q^2) \, , \tag{138}$$

where $j_\mu^c = \bar{c}\gamma_\mu c$ is the c-quark part of the quark e.m. current j_μ^{em}. Following the same derivation as in Lecture 3, we write down the dispersion relation for $\Pi^c(q^2)$ relating Im $\Pi^c(s)$ with the ratio $R_c(s)$ defined as:

$$R_c = \frac{\sigma(e^+e^- \to charm)}{\sigma(e^+e^- \to \mu^+\mu^-)} \, , \tag{139}$$

where the cross section $\sigma(e^+e^- \to charm)$ includes hadronic states with $\bar{c}c$ content produced in e^+e^-: charmonium resonances ($J/\psi, \psi', ...$), pairs of charmed hadrons, etc. In addition to the perturbative c-quark loop diagrams in Fig 14, one has to include also the contribution of the diagrams shown in Fig 20(a), (b), (c), with c quarks emitting vacuum gluons. (Remember that the nonperturbative quark condensate for the heavy c quark is absent.) The resulting relation has the following form:

$$\int\limits_{m_\psi^2}^{\infty} \frac{ds}{s-q^2} R_c(s) = \Pi^{pert}(q^2, m_c^2, \alpha_s) + \frac{\alpha_s}{\pi}\langle GG\rangle f_c(q^2, m_c^2)\,, \tag{140}$$

where m_ψ is the mass of the lowest J/ψ state in this channel. The function Π^{pert} evaluated from the massive c -quark loop diagrams has a more complicated form than in the massless case. The function f_c is the calculable short-distance part of the diagrams in Fig. 20(a), (b), (c). To achieve a better convergence at $s \to \infty$, the dispersion relation is usually differentiated n times. Note that in this case the point $q^2 = 0$ is also accessible: c quarks are still highly virtual at $q^2 = 0$ because the long-distance region starts at $q^2 \sim 4m_c^2$. The set of power moments obtained from (140) with $O(\alpha_s^2)$ accuracy has a form

$$\int\limits_{m_\psi^2}^{\infty} \frac{ds}{s^{n+1}} R^c(s) = \frac{3Q_c^2}{(4m_c^2)^n} r_n \left[1 + \alpha_s(m_c)a_n + (\alpha_s(m_c))^2 a'_n + b_n \frac{\langle \frac{\alpha_s}{\pi} GG\rangle}{(4m_c^2)^2}\right]\,, \tag{141}$$

with calculable coefficients r_n, a_n, a'_n, b_n. The natural scale for α_s in the perturbative loops is in this case the virtuality $Q \sim m_c$. The moments (141) are used to extract the gluon condensate density and, simultaneously the c quark mass, employing the experimental data on $R_c(s)$ on the l.h.s. The estimate of the condensate density obtained first in Ref. [16] is:

$$\frac{\alpha_s}{\pi}\langle GG\rangle = (330\ \text{MeV})^4 \pm 50\%\,, \tag{142}$$

again within the range of Λ_{QCD}. The value of the gluon condensate density is usually given multiplied by α_s for convenience, because this product is scale-independent.

5 RELATING QUARKS AND HADRONS: QCD SUM RULES

5.1 Introducing the method

The relation (141) obtained first in [17] is a well-known example of a *QCD sum rule*. The method developed by Shifman, Vainshtein and Zakharov [16] employs quark-current correlation functions calculating them in the spacelike region, including perturbative and condensate contributions. Consider for example the correlation function (114), with the result of QCD calculation having the form (128). Note that in the latter expression short- and long-distance effects are separated. The perturbative part $\Pi^{pert}(q^2)$ and the coefficients $C_d(q^2)$ take into account short-distance quark–gluon interactions with characteristic momenta larger than the scale Q. Both Π^{pert} and C_d are *process-dependent*, i.e., depend on the choice of the currents. On the other hand, the condensate densities absorb, in an averaged way, the long-distance interactions with momenta less than Q and are process-independent. The universality of condensates allows one to calculate correlation functions in different channels without introducing new inputs, in an almost model-independent way[11]. To obtain the sum rule, the QCD result for the correlation function is matched, via a dispersion relation, to the sum over hadronic contributions (the integral over hadronic cross section).

[11]The expansions similar to (128) represent another example of OPE in QCD (see Lecture 3.3), when a product of two currents is expanded in a set of local operators O_d. The perturbative part in this case is interpreted as a unit operator with no dimension.

A twofold use of the sum rule relations is possible. Firstly, using experimental data, one saturates the hadronic sum in the dispersion integral and extracts the universal QCD parameters: quark masses, α_s, condensate densities. One example is the $c\bar{c}$ sum rule (141) discussed above. Secondly, having fixed QCD parameters, one calculates, with a certain accuracy, the masses and decay amplitudes of the lowest hadrons entering the dispersion integral. The relation (135) obtained from the sum rule (134) can serve as an example. Note that (134) is unique, because all states in the hadronic sum except the pion and all QCD terms except the quark condensate are absent at $O(m_{u,d})$.

5.2 The ρ-meson decay constant

To demonstrate in more detail how the method works[12], let me outline the calculation of the ρ-meson decay constant, from the correlation function (114). The dispersion relation for $\Pi_{\mu\nu}$ can be written in the following form:

$$\Pi_{\mu\nu}(q) = \frac{1}{2\pi}\int\limits_{s_{min}}^{\infty}\frac{ds}{s-q^2-i\delta}\left\{\sum_{h_n}\langle 0|j_\mu|h_n\rangle\langle h_n|j_\nu(q)|0\rangle\right\}$$

$$= \int\limits_{s_{min}}^{\infty}\frac{ds}{s-q^2-i\delta}\left\{\left[\frac{f_\rho^2 m_\rho^2}{2}\delta(s-m_\rho^2)+\rho^h(s)\right](-g_{\mu\nu}q^2+q_\mu q_\nu)\right\}_{q^2=s}, \qquad (143)$$

where the ρ-meson contribution ($h_n = \rho^0$) is isolated from the sum and the hadronic matrix element is substituted:

$$\langle 0|j_\mu|\rho^0\rangle = \frac{f_\rho}{\sqrt{2}} m_\rho \epsilon_\mu^{(\rho)}\,, \qquad (144)$$

determined by the ρ decay constant f_ρ. The integrand $\rho^h(s)$ (*spectral density*) includes the sum over excited and multihadron states. We take into account the experimental fact that the ρ resonance strongly dominates in the low-energy region $2m_\pi < \sqrt{s} < 1$ GeV, so that $\rho^h(s)$ practically starts from some threshold value $s_0 \sim 1$ GeV. Switching to the invariant amplitude $\Pi(q^2)$ and using the result (128) of QCD calculation, one obtains from (143) the desired relation, a prototype of the QCD sum rule:

$$\frac{f_\rho^2 m_\rho^2}{2(m_\rho^2-q^2)} + \int\limits_{s_0}^{\infty}\frac{ds\;\rho^h(s)}{s-q^2-i\delta} = \Pi^{pert}(q^2) + \sum_{d=3,4,\ldots} C^d(q^2)\langle 0|O_d|0\rangle\,, \qquad (145)$$

which is valid at sufficiently large $|q^2|$. To proceed, one applies to both sides of this equation the *Borel transformation* defined as:

$$\hat{B}_{M^2}\Pi(q^2) = \lim_{\substack{-q^2,n\to\infty \\ -q^2/n=M^2}} \frac{(-q^2)^{(n+1)}}{n!}\left(\frac{d}{dq^2}\right)^n \Pi(q^2) \equiv \Pi(M^2)\,. \qquad (146)$$

This transformation deserves a clarifying comment. Differentiating $\Pi(q^2)$ many times in q^2, means that one is effectively approaching the long-distance region. Indeed, with an infinite amount of derivatives the function $\Pi(q^2)$ is defined at any q^2, including $q^2 > 0$. The $q^2 \to -\infty$ limit works in the opposite direction: one penetrates into the deep spacelike asymptotics. Combining two transformations in (146) at $M^2 = -q^2/n$, one fixes the virtuality scale at $O(M)$. Using the school-textbook definition of the exponent: $e^x = \lim_{n\to\infty}(1+x/n)^n$, it is an easy exercise to prove that

$$\hat{B}_{M^2}\left\{\frac{1}{m_h^2-q^2}\right\} = e^{-m_h^2/M^2}. \qquad (147)$$

[12]Reviews can be found, for example, in Refs. [15] and [18].

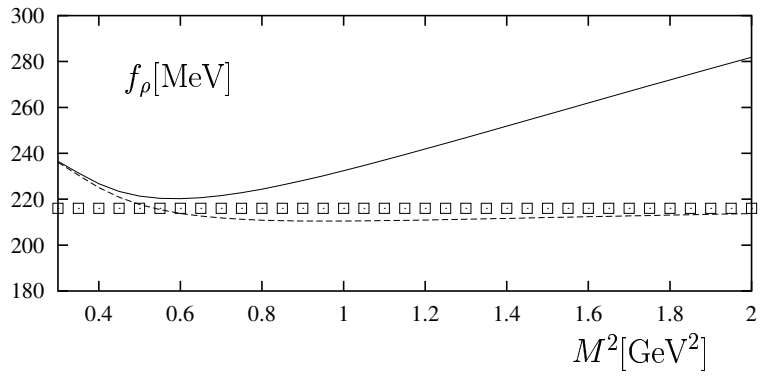


Fig. 21: The ρ meson decay constant calculated from the sum rule (148) neglecting all excited and continuum states (solid), as a function of the Borel parameter, in comparison with the experimental value (boxes). The dashed curve corresponds to an improved calculation, where the sum over excited and continuum states is estimated using quark–hadron duality with a threshold $s_0^\rho = 1.7$ GeV^2.

As a result, the Borel transformation exponentially suppresses the integral over $\rho^h(s)$ in (145) with respect to the ρ-meson term. Furthermore, after applying $\hat{B}$ to the r.h.s. of (145) the coefficients $C_d(M)$ contain powers of $1/M^2$. Hence, at large M^2 it is possible to retain only a few low-dimension condensates in the sum, e.g., at $M^2 \sim 1$ GeV^2 a reasonable approximation is to neglect all operators with $d > 6$. The explicit form of the QCD sum rule (145) after Borel transformation is [16]:

$$f_\rho^2 e^{-m_\rho^2/M^2} + \int\limits_{s_0}^{\infty} ds\ \rho^h(s) e^{-s/M^2}$$
$$= M^2\left[\frac{1}{4\pi^2}\left(1+\frac{\alpha_s(M)}{\pi}\right) + \frac{(m_u+m_d)\langle\bar{q}q\rangle}{M^4} + \frac{1}{12}\frac{\langle\frac{\alpha_s}{\pi}G^a_{\mu\nu}G^{a\,\mu\nu}\rangle}{M^4} - \frac{112\pi}{81}\frac{\alpha_s\langle\bar{q}q\rangle^2}{M^6}\right]. \qquad (148)$$

In obtaining the above relation, the four-quark vacuum densities are factorized into a product of quark condensates. The quark–gluon condensate has very small coefficient and is neglected. The running coupling α_s is taken at the scale M, i.e., at the characteristic virtuality of the loop diagrams after the Borel transformation.

Importantly, there exists a SVZ [16] region of intermediate M^2 where the ρ meson contribution alone saturates the l.h.s. of the sum rule (148).

To illustrate this statement numerically, in Fig. 21 the experimentally measured f_ρ (obtained from the $\rho^0 \to e^+e^-$ width) is compared with the same hadronic parameter calculated from the sum rule (148) where all contributions of excited and continuum states are neglected. One indeed observes a good agreement in the region $M^2 \sim 1$ GeV^2.

An important step to improve the sum rule (148) is to use the *quark–hadron duality* approximation. The perturbative contribution to the correlation function (the sum of Fig. 14 diagrams) is represented in the form of a dispersion integral split into two parts:

$$\Pi^{pert}(q^2) = \int\limits_0^{s_0} \frac{ds\ \rho^{pert}(s)}{s - q^2 - i\delta} + \int\limits_{s_0}^{\infty} \frac{ds\ \rho^{pert}(s)}{s - q^2 - i\delta}\ . \qquad (149)$$

The integral over the spectral function $\rho^h(s)$ in (145) is approximated by the second integral over the perturbative spectral density $\rho^{pert}(s)$ in (149). The latter integral is then subtracted from both parts of (145). Correspondingly (148) is modified: the l.h.s. contains only the ρ term, and, on the r.h.s., the perturbative contribution has to be multiplied by a factor $(1 - e^{-s_0/M^2})$. The numerical result obtained

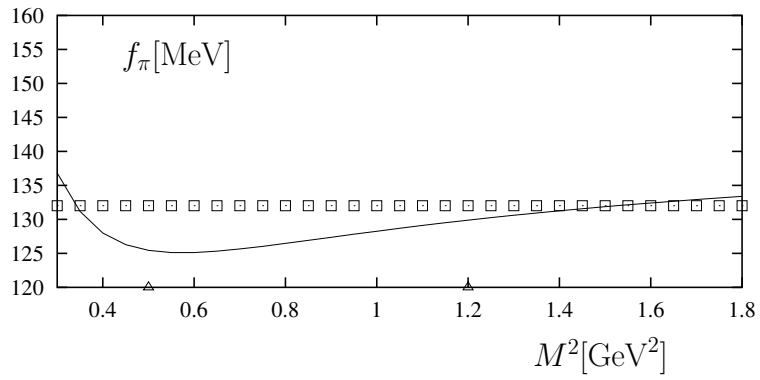


Fig. 22: The π meson decay constant calculated from the QCD sum rule [16] for the correlation function of axial-vector currents (solid line), in comparison with the experimental value (boxes). The quark–hadron duality threshold is fitted simultaneously as $s_0^\rho = 0.7$ GeV^2. The uncertainty of about 10–15% has to be added to the theoretical curve.

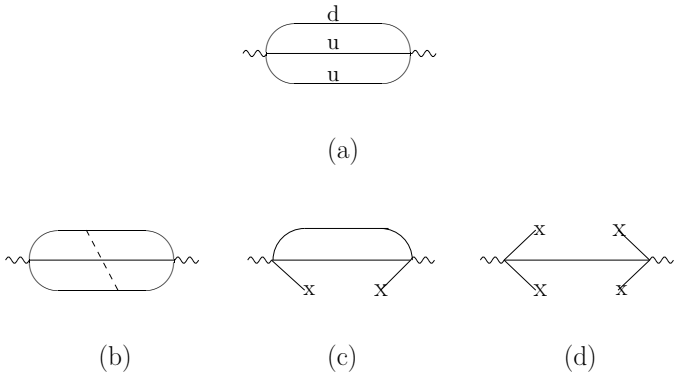


Fig. 23: Some of the diagrams contributing to the correlation function of two baryon currents: (a) the lowest-order tree-quark two-loop diagram; (b) the $O(\alpha_s)$ correction, (c),(d) quark condensate diagrams.

from the duality-improved sum rule (148) is also shown in Fig. 21. The agreement between the sum rule prediction and experiment is impressive:

$$f_\rho^{(QCDSR)} = 213 \text{ MeV} \pm (10-15)\% \tag{150}$$

whereas $f_\rho^{exp} = 216 \pm 5$ MeV [4]. The estimated theoretical uncertainty quoted in (150) is typical for QCD sum rules, reflecting the approximate nature of this method. So far we only reproduced f_ρ, using the experimental value of m_ρ. In principle, it is possible to go one step further and estimate also the ρ-mass from the sum rule. Furthermore, a very similar sum-rule analysis for the correlation function of axial currents (129) (the invariant amplitude Π_5) successfully reproduces the value of f_π (see Fig. 22).

5.3 Baryons

It is possible to extend the method of QCD sum rules to baryons [19, 20]. The idea is to construct special quark currents with baryon quantum numbers which can serve as a source of baryon production/annihilation from/to QCD vacuum. In reality, such currents hardly exist [13], but they are allowed

[13] In models of Grand Unification predicting proton decay via intermediate superheavy particles, the currents we are discussing are realized effectively in a form of localized 3-quark operators annihilating the proton.

in QCD if the colour-neutrality is obeyed. A well-known example is the Ioffe current with the nucleon quantum numbers (i.e., spin 1/2):

$$J^N(x) = \epsilon_{abc}(u^{aT}(x)\hat{C}\gamma_\mu u^b(x))\gamma_5\gamma^\mu d^c(x)\,, \tag{151}$$

where a, b, c are colour indices, $\hat{C} = \gamma_2\gamma_0$ is the charge conjugation matrix, and, for definiteness, the proton flavour content is chosen. As a next step, one constructs a correlation function

$$\Pi_N(q) = i\int d^4x e^{iqx}\langle 0|J^N(x)\bar{J}^N(0)|0\rangle\,. \tag{152}$$

The corresponding diagrams are shown in Fig. 23, including perturbative loops and vacuum condensates. They look quite different from the quark–antiquark loops, but differences concern the short-distance parts of the diagrams. The universality of condensates allows one to calculate $\Pi_N(q)$ without introducing new input parameters. The hadronic contribution contains a total sum over states produced and annihilated by the current J^N, starting from the lowest possible state, the nucleon:

$$\Pi_N(q) = \frac{\langle 0|J^N|N\rangle\langle 0|\bar{J}^N|N\rangle}{m_N^2 - q^2} + \{\text{excited resonances, multiparticle states}\}\,. \tag{153}$$

The derivation of the QCD sum rule is done along the same lines as in the previous subsection. Omitting the details, let me only mention that from this sum rule an approximate formula for the nucleon mass is obtained,

$$m_N \simeq [-(2.0)(2\pi)^2\langle 0|\bar{q}q|0\rangle(\mu = 1\text{GeV})]^{1/3}\,, \tag{154}$$

relating it to the quark-condensate density. In fact, the quark masses $m_{u,d}$ themselves generate very small corrections to this relation and are neglected. Thus, QCD sum rules provide an answer to the question that was raised in Lecture 1: almost 99% of the baryonic mass in the Universe is due to the vacuum condensates.

5.4 Quark mass determination

As already mentioned, the sum rule (141) can be used to extract the c-quark mass. Here, the role of Borel transformation is played by a simple differentiation at $q^2 = 0$, which turns out to be more useful. At low n, the moments (141) are especially convenient for m_c determination because the gluon condensate effects are small. Replacing $c \to b, \psi \to \Upsilon$, etc., one obtains analogous sum rule relations for the b quark, where the gluon condensate is much less important, being suppressed by m_b^{-4}. Recent analysis [21] of these sum rules yields $m_c(m_c) = 1.304 \pm 0.027$ GeV, $m_b(m_b) = 4.209 \pm 0.05$ GeV. Another subset of charmonium sum rules (higher moments at fixed large $q^2 < 0$) was recently employed in Ref. [22], with a prediction for m_c in agreement with the above.

The heavy-quark mass determination using sum rules is also done in a different way, employing the large n moments (141) which are less sensitive to the cross-section above the open flavour threshold. These moments, however, demand a careful treatment of Coulomb interactions between heavy quark and antiquark in the perturbative diagrams. Remember that one-gluon exchange yields Coulomb potential. Close to the threshold of heavy quark-pair production, $\sqrt{s} \simeq 2m_Q$ ($Q = b, c$) this part of the quark–antiquark interaction becomes important, and at large n the near-threshold region dominates in the perturbative coefficients a_n, a'_n in (141). A systematic treatment of this problem is possible in *nonrelativistic QCD (NRQCD)*, a specially designed effective theory obtained from QCD in the infinite heavy-quark mass limit (for a review see, for example, Ref. [23]).

Equally well, QCD sum rules allow one to estimate the masses of light u, d, s quarks. To give only one typical example from the vast literature, let me refer to the recent analysis [24], based on the correlation function for derivatives of the s-flavoured vector current $j_{s\,\mu} = \bar{s}\gamma_\mu q$, ($q = u, d$):

$$\Pi^s(q) = i\int d^4x e^{iqx}\langle 0|T\{\partial_\mu j_s^\mu(x)\partial_\nu j_s^{\dagger\nu}(0)\}|0\rangle\,. \tag{155}$$

The QCD answer for Π^s is proportional to $(m_s - m_q)^2 \simeq m_s^2$ (due to $\partial_\mu j_s^\mu = (m_s - m_q)\bar{s}q$) making this correlator very sensitive to m_s. One calculates the usual set of diagrams shown in Figs. 14 and 20, where the quark lines are now s and q. Furthermore, the recent progress in the multiloop QCD calculations allows one to reach the $O(\alpha_s^3)$ accuracy in the perturbative part of Π^s and to include also $O(\alpha_s)$ corrections to the condensate contributions. The hadronic spectral density $\rho_K(s)$ for this correlator is saturated by $J^P = 0^+$ states with s-flavour, e.g., $K\pi$ states with $L = 0$. The sum rule has the form

$$\int ds \rho_K(s) e^{-s/M^2} = (m_s - m_{u,d})^2 \left[\Pi^{pert}(M) + \frac{C_4^K}{M^4} + \frac{C_6^K}{M^6} \right] , \tag{156}$$

Data of kaon S-wave scattering on π, η, η' were used to reproduce $\rho^K(s)$. The resulting prediction [24] for the mass is

$$m_s(2\text{ GeV}) = 99 \pm 16\text{ MeV}. \tag{157}$$

The ratios of the light (u, d, s) quark masses can be predicted in QCD from the relations for pions and kaons, similar to (135). A systematic derivation is done employing *chiral perturbation theory*, an effective theory obtained from QCD in the a low-energy limit, using instead of quarks and gluons, the pion and kaon degrees of freedom (for a review see Ref. [25]). The result is [26]:

$$\frac{m_u}{m_d} = 0.553 \pm 0.043, \quad \frac{m_s}{m_d} = 18.9 \pm 0.8, \quad \frac{2m_s}{m_u + m_d} = 24.4 \pm 1.5\,. \tag{158}$$

From the above ratios and the value (157) one obtains $m_u(2\text{ GeV}) = 2.9 \pm 0.6$ MeV and $m_d(2\text{ GeV}) = 5.2 \pm 0.9$ MeV.

We see that QCD sum rules are extremely useful for the quark mass determination. The m_q values extracted from sum rules are included, together with the lattice determinations and results of other methods, in the world-average intervals in Ref. [4] presented in (31).

5.5 Calculation of the B-meson decay constant

In B-meson decays the CKM parameters of the Standard Model are inseparable from hadronic matrix elements. Hence, without QCD calculation of these matrix elements with an estimated accuracy, it is impossible to use experimental data on B decays for extracting the Standard Model parameters and for detecting/constraining new physics effects. Currently, lattice QCD provides many hadronic parameters for B physics, with a continuously improving accuracy. QCD sum rules represent another actively used working tool. With condensates and quark masses determined from a set of experimentally proven sum rules for light-quark and heavy quarkonium systems, one has a real possibility to assess the theoretical accuracy of the sum rule predictions by varying the input within allowed intervals.

One of the most important applications of QCD sum rules is the determination of the B-meson decay constant f_B defined via the matrix element

$$m_b \langle 0 \mid \bar{q} i\gamma_5 b \mid B \rangle = m_B^2 f_B\ , \tag{159}$$

($q = u, d, s$). Note that f_B multiplied by V_{ub} determines the width of leptonic B decays, such as $B^- \to \tau^- \bar{\nu}_\tau$. To calculate f_B from QCD sum rules, one usually employs the correlation function:

$$\Pi_5^{(B)}(q) = i \int d^4x e^{iqx} \langle 0 | T\{ m_b \bar{q}(x) i\gamma_5 b(x), m_b \bar{b}(0) i\gamma_5 q(0) \} | 0 \rangle\ , \tag{160}$$

so that the lowest B meson term in the hadronic sum for the above correlation function contains f_B:

$$\Pi_5^{(B)}(q) = \frac{m_B^4 f_B^2}{m_B^2 - q^2} + \ldots\ . \tag{161}$$

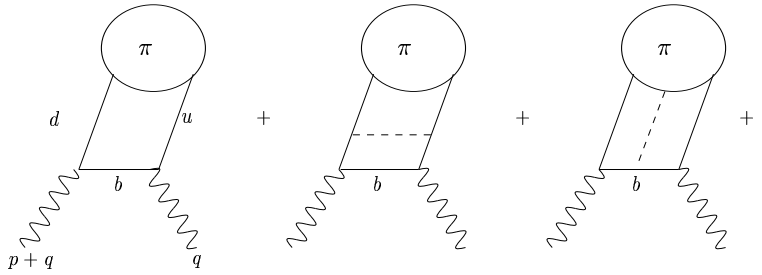


Fig. 24: Diagrams contributing to the correlation function (163), from left to right: the leading-order, $O(\alpha_s)$ correction, soft gluon. The blob with π denotes the pion distribution amplitude.

To obtain the sum rule, one needs to calculate $\Pi_5^{(B)}(q)$ from the perturbative and condensate diagrams in Figs. 14 and 20, with γ_5 vertices emitting and absorbing b and q lines. One, therefore needs an input value for m_b and m_s ($m_{u,d}$ can safely be neglected). These values are taken from the analyses overviewed in the previous subsection. The sum rule has a form similar to the one for f_ρ. Naturally, the expressions for the perturbative part and for the coefficients C_d are completely different. Also the hierarchy of contributions in the heavy-light correlation function differs from the light-quark case. Now the quark condensate becomes very important being proportional to $m_b\langle\bar{q}q\rangle$. The recent updates of the sum rule for f_B obtained in Refs. [27,28] take into account the $O(\alpha_s^2)$ corrections to the heavy-light loop calculated in Ref. [29]. The numerical prediction of the sum rule, taking $m_b(m_b) = 4.21 \pm 0.05$ GeV, is $f_B = 210 \pm 19$ MeV and $f_{B_s} = 244 \pm 21$ MeV [27], in good agreement with the most recent lattice QCD determinations.

5.6 Light-cone sum rules and $B \to \pi$ form factor

To complete our brief survey of QCD sum rules, let me introduce one important version of this method, the light-cone sum rules (LCSR) [30, 31] used to calculate various hadronic amplitudes relevant for exclusive processes. In the following, we consider the application of LCSR to the $B \to \pi$ transition amplitude (see Fig. 9). The latter is determined by the hadronic matrix element

$$< \pi^+(p)|\bar{u}\gamma_\mu b|B(p+q) >= 2f_{B\pi}^+(q^2)p_\mu + \left[f_{B\pi}^+(q^2) + f_{B\pi}^-(q^2)\right] q_\mu \,, \tag{162}$$

generated by the $b \to u$ weak current (67). Owing to spin-parity conservation only the vector part of the current contributes. There are two independent 4-momenta p and q, and one independent invariant q^2, the momentum transfer squared. The initial and final mesons are on shell, $p^2 = m_\pi^2$ and $(p+q)^2 = m_B^2$. It is quite obvious that one needs two invariant functions of q^2, the *form factors* $f_{B\pi}^+(q^2)$ and $f_{B\pi}^-(q^2)$, to parametrize this matrix element. Only one form factor $f_{B\pi}^+$ is interesting, the other one is kinematically suppressed in the measurable $B \to \pi l \nu_l$ semileptonic decay rate ($l = e, \mu$).

To derive LCSR for $f_{B\pi}^+(q^2)$ one uses a new type of correlation function, which itself represents a hadronic matrix element. It is constructed from the product of the weak $\bar{u}\gamma_\mu b$ current and the current $m_b\bar{b}i\gamma_5 d$ used to generate B in (159). The currents are taken at two different 4-points and sandwiched between vacuum and the one-pion state. The formal definition of this correlation function reads:

$$\begin{aligned} F_\mu(q,p) = i\int d^4x e^{iqx}\langle\pi^+(p) \mid T\{\bar{u}(x)\gamma_\mu b(x), m_b\bar{b}(0)i\gamma_5 d(0)\} \mid 0\rangle \\ = F(q^2,(p+q)^2)p_\mu + \tilde{F}(q^2,(p+q)^2)q_\mu \,, \end{aligned} \tag{163}$$

where it is sufficient to consider only one invariant amplitude F. Note that we now have two kinematical invariants, q^2 and $(p+q)^2$, still $p^2 = m_\pi^2$.

Diagrammatically, the correlation function (163) is represented in Fig. 24. At $q^2 < 0$ and spacelike $(p+q)^2 < 0$ very similar diagrams describe the process $\gamma^*\gamma^* \to \pi^0$, the one-pion production by two virtual photons via e.m. currents[14]. One only has to replace all quarks in the diagrams in Fig. 24 by either u or d quarks. Both objects, the heavy-light correlation function (163) and the $\gamma^*\gamma^* \to \pi^0$ amplitude, contain one virtual quark propagating between vertices and a quark–antiquark pair which is emitted at points x and 0 and converted into a real pion state. At large spacelike external momenta $|(p+q)^2|, |q^2| \gg \Lambda^2_{QCD}$ the space–time interval $x^2 \simeq 0$ approaches the light-cone. Hence, the virtual quark in both $\langle$two-currents $\to$ pion$\rangle$ amplitudes propagates at short distances allowing a perturbative QCD description. The calculable short-distance parts are process-dependent. In the case of the correlation function (163) we have a virtual b quark propagating between vertices of flavour-changing currents, whereas in the $\gamma^*\gamma^* \to \pi^0$ amplitude the light quark propagates between e.m. vertices. The long-distance part in both cases is, however, the same vacuum-to-pion matrix element of light quark and antiquark emitted at points x and 0:

$$\langle \pi(p) \mid \bar{q_1}(x)\Gamma_a q_2(0) \mid 0\rangle\,, \tag{164}$$

where in one case $q_1 = u$, $q_2 = d$, $\pi = \pi^+$ and in the other case $q_1 = u(d)$, $q_2 = u(d)$, $\pi = \pi^0$. Owing to isospin symmetry, the difference between these two configurations is indeed very small. In order to treat the diagram with one extra gluon entering the pion, one has to introduce an additional quark–antiquark–gluon matrix elements of the type

$$\langle \pi(p) \mid \bar{u}(x) g_s G^{\mu\nu}(y)\Gamma_b d(0) \mid 0\rangle\,, \tag{165}$$

where $x^2 \sim y^2 \sim (x-y)^2 \to 0$. In the above $\Gamma_{a,b}$ denote certain combinations of Dirac matrices.

Having separated short and long distances, one is able to calculate the correlation function (163) in a form of *light-cone OPE*, where the short-distance part (the virtual b quark propagator plus gluon corrections) is multiplied by a long-distance part, the universal matrix elements such as (164), (165). The latter can be parametrized in terms of *light-cone distribution amplitudes* of the pion [32], The most important of them is defined by

$$\langle \pi(p)|\bar{u}(x)\gamma_\mu\gamma_5 d(0)|0\rangle = -ip_\mu f_\pi \int_0^1 du\, e^{iupx}\varphi_\pi(u,\mu)\,, \tag{166}$$

where μ is a characteristic momentum scale, determined by the average x^2 in the correlation function. In the above definition, u and $1-u$ are the fractions of the pion momentum p carried by the constituent quark and antiquark, in the approximation where one neglects the transverse momenta of the constituents with respect to the longitudinal constituents. The leading-order answer for F obtained from the first diagram in Fig. 24 using (166) is quite simple

$$F(q^2,(p+q)^2) = m_b f_\pi \int_0^1 \frac{du\;\varphi_\pi(u,\mu)}{m_b^2 - (q+up)^2}\,. \tag{167}$$

The pion DA $\varphi_\pi(u)$ plays here the same role of nonperturbative input as the condensates in the conventional QCD sum rules considered above. Asymptotically, that is at $\mu \to \infty$, QCD perturbation theory implies $\varphi_\pi(u,\infty) = 6u(1-u)$. However, at the physical scale $\mu \sim m_b$, at which the OPE is applied to the correlation function (163), nonasymptotic effects are to be expected, which we shall not discuss for the sake of brevity. Importantly, there is a power hierarchy of different contributions stemming from diagrams in Fig. 24 determined by the large scale in the correlation function. This scale is given by the virtuality of the b quark, i.e., by the quantity which stands in the denominator of the b -quark propagator: $m_b^2 - (q+up)^2 = m_b^2 - (p+q)^2 u - q^2(1-u)$. Importantly, this quantity remains large when q^2 is positive (timelike) but not very large, $q^2 \ll m_b^2$, allowing one to penetrate into the lower part of the kinematical region $0 < q^2 < (m_B - m_\pi)^2$ of the $B \to \pi$ transition. An example of a subleading contribution

[14]This process is experimentally accessible in $e^+e^- \to e^+e^-\pi^0$ two-photon (double-tagged) collisions.

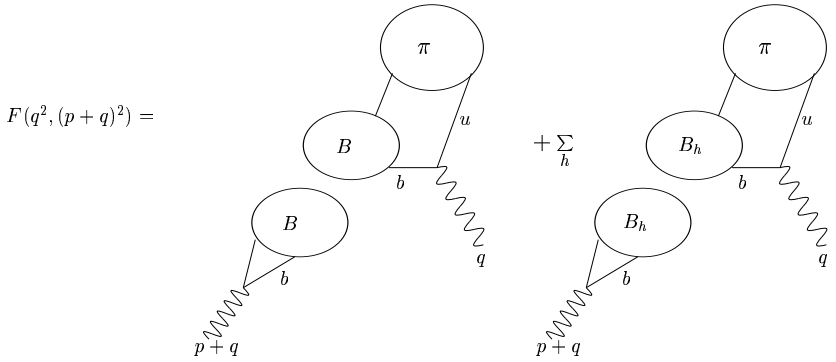


Fig. 25: A schematic dispersion relation.

is the diagram with an additional gluon entering pion DA (the third one in Fig. 24). It has two powers of inverse scale and is suppressed with respect to the leading-order diagram. Over recent years $O(\alpha_s)$ corrections and quark–antiquark–gluon contributions have been calculated to improve the result (167).

Having evaluated F as a function of q^2 and $(p+q)^2$, we are still half-way from the final sum rule. The next important step is writing a dispersion relation in the variable $(p+q)^2$, the external momentum of the current with B meson quantum numbers. The diagrammatical representation of the dispersion relation is shown in Fig. 25. It contains the same set of hadronic states as in the sum rule for f_B, starting from the B meson ground state. Using (159) and (8) one obtains the hadronic representation we need:

$$F(q^2,(p+q)^2) = \frac{2m_B^2 f_B f_B^+(p^2)}{m_b(m_B^2-(p+q)^2)} + \dots \ . \tag{168}$$

The ellipses in the above denote the contributions from the excited B and from continuum states. Equating the QCD result (167) in the region of validity at $(p+q)^2 < 0$ with the dispersion relation (168), one obtains a raw sum rule relation for $f_B f_B^+(p^2)$. The rest of the calculation follows the usual QCD sum rule procedure: Borel transformation in $(p+q)^2$ and subtraction of the contribution from higher states invoking quark–hadron duality. One finally arrives at an expression for the desired form factor, the dominant term of which obtained directly from (167) is given by

$$f_{B\pi}^+(q^2) = \frac{f_\pi m_b^2}{2f_B m_B^2}\int_\Delta^1 \frac{du}{u}(\varphi_\pi(u,\mu_b)+\dots)\exp\left(\frac{m_B^2}{M^2}-\frac{m_b^2-q^2(1-u)}{uM^2}\right) . \tag{169}$$

Here, M is the Borel mass parameter, and the scale μ_b is of the order of the characteristic virtuality of the correlation function, $\mu_b^2 = m_B^2 - m_b^2$. The integration limit $\Delta = (m_b^2-p^2)/(s_0-p^2)$ depends on the effective threshold s_0^B above which the contribution from higher states to the dispersion relation (168) is cancelled against the corresponding piece in the QCD representation (167). The parameters f_B, s_0^B are usually taken from the sum rule for f_B considered in the previous subsection. The most recent predictions for $f_{B\pi}^+(q^2)$ obtained from LCSR [33] were used to extract $|V_{ub}|$ from the measurements of the exclusive semileptonic width at B factories [34], using the formula for the decay rate:

$$\frac{d\Gamma(B\to\pi l\bar\nu)}{dq^2} = \frac{G^2|V_{ub}|^2}{24\pi^3}(E_\pi^2-m_\pi^2)^{3/2}[f_{B\pi}^+(q^2)]^2 , \tag{170}$$

where E_π is the pion energy in the B meson rest frame.

6 CONCLUSIONS

QCD has a thirty-year history and embraces many approaches, some of them developed quite independently from the others. Owing to self-interactions of gluons, this theory has an extremely rich dynamics,

combining asymptotic freedom at short distances with the self-emerging energy scale Λ_{QCD} and confinement at long distances. Accordingly, QCD has two different phases: the perturbative one responsible for the quark–gluon processes at large momentum transfers, and the nonperturbative one where the only observable states are hadrons formed by confined quarks and gluons. Yet there is no complete analytical solution for the hadronic phase of QCD. One has to rely on approximations: either numerical (QCD on the lattice) or analytical (QCD sum rules). In addition, several effective theories corresponding to different limits of QCD and exploiting the rich symmetry pattern of the theory are successfully used. In the overview of QCD given in these lectures I tried to emphasize the importance of the hadronic aspects of QCD, where the most nontrivial phenomena and challenging problems are accumulated.

ACKNOWLEDGEMENTS

I would like to thank the organizers of the School for inviting me to give these lectures and for an enjoyable meeting at Tsakhkadzor. I benefited from the help of discussion leaders. My special thanks are to the students of the School for their stimulating interest and for many questions and comments. This work was partially supported by the German Ministry for Education and Research (BMBF).

REFERENCES

[1] I. Aitchison, Lectures at this School.

[2] Kostandin Erznkazi, *The Book of Morning Light* (in Armenian), (Editions 'S. Grokh', Yerevan, 1961).

[3] DELPHI Collaboration, Events on the Z peak displayed at http://delphiwww.cern.ch/delfigs/events/z0ps/z0maxen.html .

[4] K. Hagiwara *et al.* (Particle Data Group Collaboration), Phys. Rev. D **66** (2002) 010001.

[5] S. R. Sharpe, Lectures given at Theoretical Advanced Study Institute, Boulder, CO, 1994: CP violation and the limits of the Standard Model, hep-ph/9412243;
A. S. Kronfeld, *in* Shifman, M. (ed.), *At the Frontier of Particle Physics*, vol. 4, Handbook of QCD (World Scientific, Singapore, 2002), pp. 2411–2477; hep-lat/0205021.

[6] R. Fleischer, Lectures at this School.

[7] L. B. Okun, *Weak Interactions of Elementary Particles* (Pergamon Press, Oxford, 1963).

[8] M. A. Shifman, *in* Boulder TASI 95:409-514, hep-ph/9510377.

[9] T. Mannel, hep-ph/9611411;
A. F. Falk, hep-ph/9812217.

[10] T. Nakano *et al.* (LEPS Collaboration), Phys. Rev. Lett. **91** (2003) 012002, hep-ex/0301020;
S. Stepanyan *et al.* (CLAS Collaboration), Phys. Rev. Lett. **91** (2003) 252001, hep-ex/0307018;
V. V. Barmin *et al.* (DIANA Collaboration), Phys. Atom. Nucl. **66** (2003) 1715, hep-ex/0304040.

[11] T. Sjostrand, L. Lonnblad, S. Mrenna and P. Skands, hep-ph/0308153.

[12] A. Pich, hep-ph/0001118.

[13] G. Altarelli, Lectures given at the *2001 European School of High-Energy Physics*, Beatenberg, Switzerland, (CERN-2002-002) Geneva, 2002, pp. 65–101; e-Print Archive: hep-ph/0204179.

[14] D. Diakonov, Prog. Part. Nucl. Phys. **51** (2003) 173, hep-ph/0212026.

[15] P. Colangelo and A. Khodjamirian, *in* Shifman, M. (ed.), *At the Frontier of Particle Physic*, vol. 3, Handbook of QCD (World Scientific, Singapore, 2002), pp. 1495–1576; hep-ph/0010175.

[16] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, Nucl. Phys. B **147** (1979) 385, 448.

[17] V. A. Novikov, L. B. Okun, M. A. Shifman, A. I. Vainshtein, M. B. Voloshin and V. I. Zakharov, Phys. Rev. Lett. **38** (1977) 626 [Erratum ibid. **38** (1977) 791]; Phys. Rep. **41** (1978) 1.

[18] M. Shifman, Prog. Theor. Phys. Suppl. **131** (1998) 1; hep-ph/9802214;
E. de Rafael, Lectures at *Les Houches Summer School*, Session 68, Les Houches, France (1997), hep-ph/9802448.

[19] B. L. Ioffe, Nucl. Phys. B **188** (1981) 317 [Erratum ibid. B **191** (1981) 591]; Z. Phys. C **18** (1983) 67.

[20] Y. Chung, H. G. Dosch, M. Kremer and D. Schall, Phys. Lett. B **102** (1981) 175; Nucl. Phys. B **197** (1982) 55.

[21] J. H. Kuhn and M. Steinhauser, Nucl. Phys. B **619** (2001) 588.

[22] B. L. Ioffe and K. N. Zyablyuk, hep-ph/0207183.

[23] A. H. Hoang, *in* Shifman, M. (ed.), *At the Frontier of Particle Physics*, vol. 4, (World Scientific, Singapore, 2002) pp. 2215–2331; hep-ph/0204299.

[24] M. Jamin, J. A. Oller and A. Pich, Eur. Phys. J. C **24** (2002) 237.

[25] H. Leutwyler, Lectures given at *14th Summer School on Understanding the Structure of Hadrons* (HADRONS 01), Prague, Czech Republic, 9–13 July 2001; hep-ph/0212325.

[26] H. Leutwyler, Phys. Lett. B **378** (1996) 313.

[27] M. Jamin and B. O. Lange, Phys. Rev. D **65** (2002) 056005.

[28] A. A. Penin and M. Steinhauser, Phys. Rev. D **65** (2002) 054006, hep-ph/0108110.

[29] K. G. Chetyrkin and M. Steinhauser, Eur. Phys. J. C **21** (2001) 319.

[30] I. I. Balitsky, V. M. Braun and A. V. Kolesnichenko, Nucl. Phys. B **312** (1989) 509;
V. M. Braun and I. E. Filyanov, Z. Phys. C **44** (1989) 157.

[31] V. L. Chernyak and I. R. Zhitnitsky, Nucl. Phys. B **345** (1990) 137.

[32] G. P. Lepage and S. J. Brodsky, Phys. Lett. B **87** (1979) 359; Phys. Rev. D **22** (1980) 2157;
A. V. Efremov and A. V. Radyushkin, Phys. Lett. B **94** (1980) 245; Theor. Math. Phys. **42** (1980) 97;
V. L. Chernyak and A. R. Zhitnitsky, JETP Lett. **25** (1977) 510; Sov. J. Nucl. Phys. **31** (1980) 544.

[33] A. Khodjamirian, R. Ruckl, S. Weinzierl, C. W. Winhart and O. Yakovlev, Phys. Rev. D **62** (2000) 114002;
P. Ball and R. Zwicky, JHEP **0110** (2001) 019.

[34] S. B. Athar *et al.* (CLEO Collaboration), Phys. Rev. D **68** (2003) 072003; hep-ex/0304019.

ON THE THERMODYNAMICS OF INELASTIC HADRON PROCESSES

J. Manjavidze, IP, Tbilisi, Georgia and JINR, Dubna, Russia
A. Sissakian, JINR, Dubna, Russia

Abstract

We discuss the S-matrix interpretation of thermodynamics for the multiple production process. Two possible boundary conditions are considered. One of them is usual in a field theory vacuum boundary condition. The corresponding thermodynamics formalism can be used in particle physics. Another type assumes that the system under consideration is in the environment of black-body radiation. The latter leads to canonical thermodynamics. The comparison with Schwinger–Keldysh real-time finite-temperature field theory and with the non-stationary statistical operator approach of Zubarev are considered. The range of applicability of the finite-temperature description of the multiple production process is shown.

1 INTRODUCTION

We shall discuss the conditions of the thermodynamic description of multiple production processes. It will be shown that in definite conditions one can actually adopt the formalism of equilibrium thermodynamics to multiple production processes: But the formulae of thermodynamics must be modified.

We hope that the questions will be of interest not only for experts in particle physics and not only for theorists. Special attention will be paid to the qualitative side of the question. The technical details are given in the Appendices.

A few words of history. The first experimental observation of a multiple production process in cosmic rays was made by Skobeltzyn in 1927. It became clear after Powell's discovery in 1947 that the cosmic rays consisted of π-mesons. But even now, after half a century of development, there is no quantitative understanding of the multiple production phenomenon. Why? We shall try to give an answer.

So, we want to understand: can one use the language of thermodynamics for multiple production processes? We think this question needs a special discussion. First of all, we are trying to describe the process of dissipation of kinetic energy into the mass of particles and this process proceeds in a vacuum, i.e. it can freely evolve. It must be noted also that we can measure only a restricted set of parameters: the three-dimensional momentum, energy, charge and spin of the mass-shell particle; the number of particles (with definite accuracy); and, finally, the number of discussed interactions in the given coordinate frame. This is all that we know! We will see what it will do for the thermodynamical description.

The idea to use thermodynamics was intensively worked out after the pioneering work of Fermi and Landau [1]. There have also been a number of modern attempts to use some properties of thermodynamics especially for high-energy ion inelastic collisions [2]. The attempts show a good agreement. But there is agreement only in particulars instants.

The experiment shows that the immediate use of thermodynamics leads to dubious predictions. For instance, the Fermi–Landau model predicts that the hadron mean multiplicity is $\sim E$, where E is the CM total energy. The experiment gives

$$\bar{n}(E) = O(\ln E) << E. \tag{1}$$

Therefore, the distinction is too large at high energies, see Fig. 1.

Nevertheless, why is thermodynamics so attractive to particle physics? One can start from the remark that both quantum and thermodynamical theories have a probabilistic background [4]. It is natural for this reason that they use such similar notions as probability, ground state, correlation functions, distribution functions (also inclusive), phase transitions, Higgs phenomena and so on.

But there is also an important property peculiar to thermodynamics. Namely, the possibility to describe a huge number of particles (the Avogadro number is $\sim 10^{24}$) with only a few parameters. Temperature and chemical potential are the most popular among them.

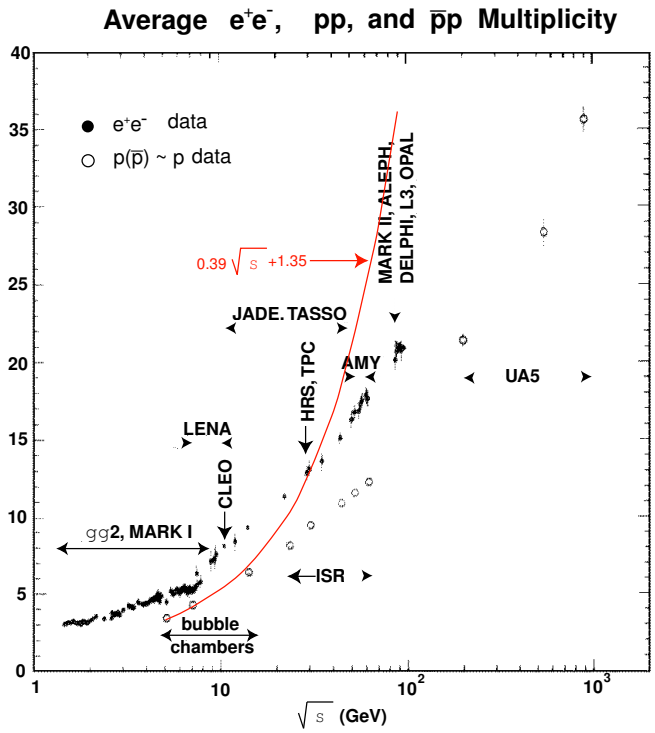


Fig. 1: Mean multiplicity [3]. The line: $\bar{n}(E) = 1.35 + 0.39\sqrt{s}$.

The question is that the amplitude of the production of n particles depends on $(3n-4)$ variables. The mean multiplicity $\bar{n}(E) \sim 100$ in the modern experiments and, therefore, even if the probabilistic character of measurements is not taken into account, there are too many degrees of freedom to be describable analytically. One may consider the integral quantities, the total cross section, or one particle differential (inclusive) cross sections. But in this case we can not be sure that the important physical properties are not lost because of integration.

There is also one more problem in hadron physics. It is connected to the high symmetry of underlying field theory. It is assumed that the future theory of hadrons will be constructed on the basis of the non-Abelian gauge theory of Yang and Mills. The reason of such a hope is based on asymptotic freedom. The observable consequence of it is Bjorken scaling which appears if the quarks are free at small distances. Another prediction is high-transverse-momentum jets.

Both phenomena, as stated above, are a short-distance effect. At these distances the role of symmetry as a source of long-range constraints is not high and one can use 'perturbative QCD' which has approximately the same structure as QED. But we will consider hadron multiple production phenomena where apparently the role of symmetries is high. The latter follows from the comparable smallness of

the hadron mean multiplicity, see Eq. (1).

Indeed, it is natural to consider multiplicity as the measure of incident energy dissipation. Then the symmetry through the conservation laws, including the 'hidden', must prevent the dissipation process and must lead to Eq. (1). Fermi, Pasta and Uhlam (FPU) were the first to observe this phenomenon [5]. They were confused when observing the negative result of the Monte Carlo experiment: the local perturbation of the one-dimensional nonlinear chain of small massive balls had not distributed over the chain degrees of freedom uniformly. The theoretical explanation of the FPU effect was given in the seventies by Zabuski, Kruskal and Zakharov [6]. It was shown (Zabuski and Kruskal) that the perturbation in the FPU chain produces the soliton. It was shown also that the FPU chain hides the high symmetry (Zakharov).

One may conclude when observing the low value of hadron mean multiplicity that the role of non-Abelian gauge symmetry is high, $\bar{n}(E) \ll E$, but can not be neglected, $\bar{n}(E) \gg 1$. Therefore, the highly unstable system is produced in high-energy hadron interactions. For example, the fluctuations over the rapidity are large, see Fig. 2.

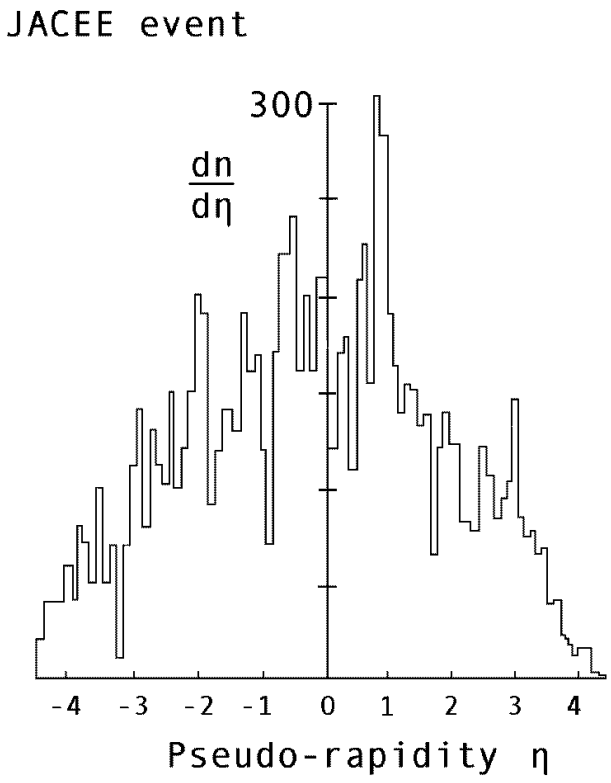


Fig. 2: Pseudo-rapidity fluctuations in the event-by-event analysis.

It must be noted that the Fermi–Landau model ignores the influence of symmetry assuming that the final state of the inelastic collision is equilibrium and the energy spreads over all degrees of freedom uniformly. At the same time, following Fermi and Landau, if one considers the very high multiplicity (VHM) final state, when $n \to n_{\max} \sim E$, then such a state will be in equilibrium and the symmetry would not play an important role during the process of dissipation [7]. This interesting idea is examined experimentally in the various experiments of CERN, FNAL, RHIC and Protvino.

It will be useful to have in mind the following toy model offered at the very beginning of the 20th century by a couple, P. and T. Ehrenfest, to explain in simple terms the Boltzmann interpretation of the irreversibility phenomenon in statistics. The model is extremely simple and fruitful [4]. It considers two

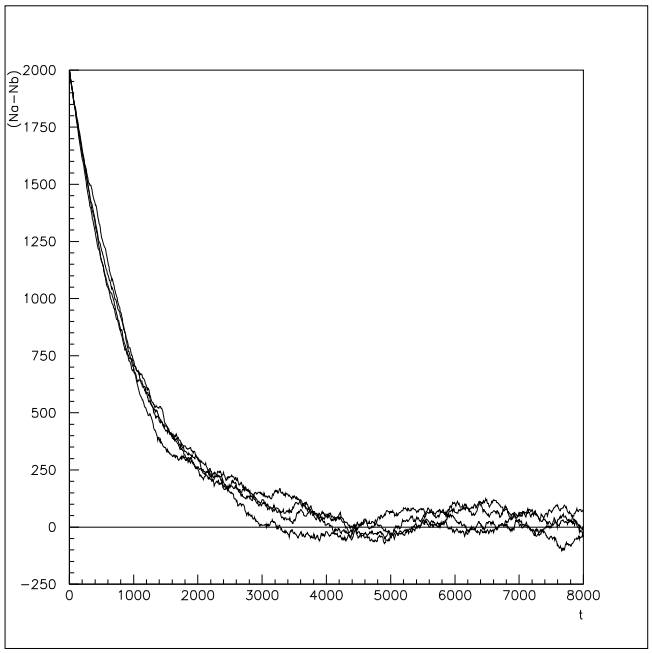


Fig. 3: Ehrenfest model description of non-equilibrium flow. Four Monte Carlo simulations are shown; t is the number of permutations; $N_a\,(N_b)$ is the number of balls in the $a(b)$ box; $N_b = 0$ at $t = 0$. The absence of fluctuation is noticeable for $t < 2000$: the fluctuation rate increases near to equilibrium.

boxes with $2N$ numerated balls. Choosing number $l = 1, 2, ..., 2N$ $randomly$ one must take the ball with the label l from the box and put it in another one. Starting from the highly 'nonequilibrium' state with all the balls in one box, the number of balls in the boxes tends to equalize, Fig. 3.

So, there is an irreversible flow towards a preferable (equilibrium) state. One can hope [4] that this model reflects a physical reality of nonequilibrium processes. If the initial state is very far from equilibrium, a theory of such processes should be sufficiently simple. Just this situation can be realized in the case of VHM processes. In the ordinary no-bias inelastic hadron experiments we have the highly fluctuating system, see Fig. 3 for $t > 2000$, i.e. when the number of balls becomes equal in both boxes.

Therefore, the equilibrium noticeably simplifies the situation since in this case the whole system can be described well by the mean value of the measurement parameters. However, to have equilibrium, it is necessary and sufficient to have Gaussian fluctuations in the vicinity of the corresponding mean value. In this case the temperature T, which coincides with the mean value of energy, would be a 'good' parameter. Exactly this supposition will be discussed.

Notice that it is hard to assume that in the no-bias inelastic hadron reactions the energy is distributed uniformly. Indeed, the energy spectra show rather the approximate Feynman scaling law [8]: the inclusive spectra over the energy ε are rather $\sim d\varepsilon/\varepsilon$, see Fig. 4. In any case, the spectra of Fig. 4 are not the Gaussian distributions.

In addition, the experimental proof of the Gaussian fluctuations is a hard task. Much more useful is to check the relaxation of correlations [7] since it signifies the situation when the system becomes 'calm'. For this purpose we will construct the S-matrix interpretation of thermodynamics.

2 S-MATRIX INTERPRETATION OF THERMODYNAMICS

2.1 S-matrix approach and thermodynamics

For the particle production process we are dealing with a non-equilibrium process. To formulate the corresponding theory one must introduce the notion of $mechanical$ perturbation. It is considered as the

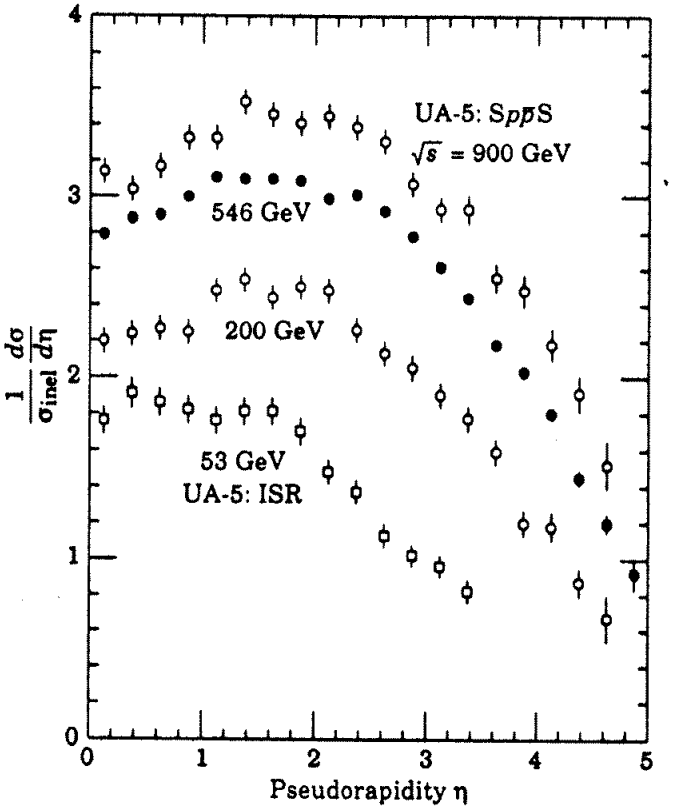


Fig. 4: Rapidity distribution.

additional energy of the controlled environment to the given Hamiltonian of the system. But at the same time there are $thermal$ perturbations. They can not be included in the Hamiltonian since they are the result of averaging over the existing degrees of freedom. Therefore, the question of the thermalization $process$ is open in the canonical formalism [9].

In our case the temperature is defined as the mean energy of produced particles. In this case it is not necessary to consider the thermalization. This process is 'hidden' in the box where the interactions were performed. This position is correct since the local theory is considered. This means that the interaction field's energy is completely transformed into the energy of produced particles. Some of it 'thermalizes', i.e. is spread as the kinetic energy, and the other part is transformed into the mass of particles.

Following Gibbs, see for example Ref. [10], we must introduce the density of states in the phase space, $d\Gamma$. The corresponding unnormalized probability is defined by the integral

$$\rho = \int d\Gamma. \tag{2}$$

The probability realized in the phase space domain Ω is defined by the ratio

$$W(\Omega) = \frac{\int_\Omega d\Gamma}{\int d\Gamma}. \tag{3}$$

It must be noted that Ω sometimes does not have a definite boundary. Then, to define Ω correctly, it is useful to introduce the 'probe functions', e.g. (α, z) [11]. Therefore, we will calculate the so-called generating functional, $\rho(\alpha, z)$.

We will distinguish the initial and final states. This distinction follows from the wave nature of our quantum process. It must be assumed that the wave should disappear on the infinite hypersurface

σ_∞ [12]. At the same time the measurable asymptotic states must be the eigenfunctions of the energy and momentum operators [13]. These two conditions can be satisfied introducing the $i\varepsilon$ prescription of Feynman. Therefore, our wave processes can not be time reversible.

We will consider the m-into-n particle transition amplitude, see Fig. 5. Then the density of state is equal to the Lorentz-covariant phase space element

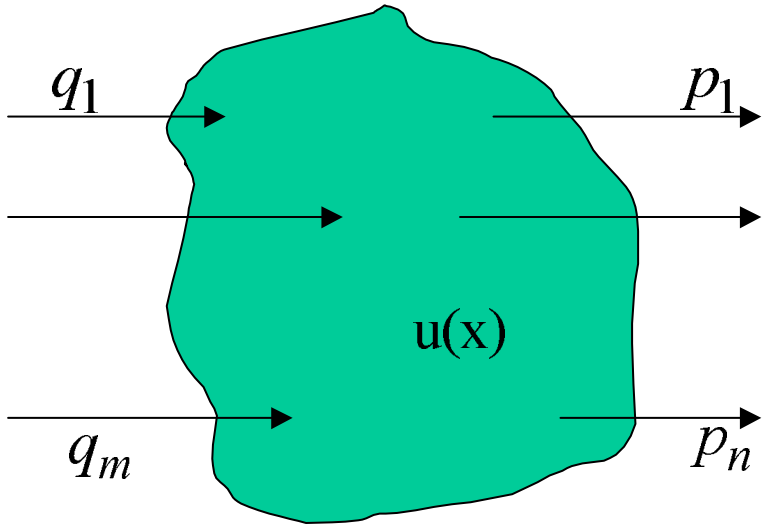


Fig. 5: The amplitude a_{mn}. The plane wave $e^{-iq_kx_k}$, $k = 1, 2, ..., m$, is associated to each incoming particle and $e^{ip_ky_k}$, $k = 1, 2, ..., n$, to each outgoing one. It is supposed that $q_k^2 = p_k^2 = \mu^2$. The integration over all configurations of $u(x)$ must be performed.

$$d\omega_m(q) = \prod_{i=1}^{m} \frac{dq_i}{(2\pi)^3 2\varepsilon_i \; m!}, \quad \varepsilon_i = +\sqrt{q_i^2 + m^2}.$$

The factor $1/m!$ is the consequence of particle identity.

It is necessary to take into account the condition that the final state can not be free from the initial one. The density of probability $|a_{m,n}(q,p)|^2$, where $a_{m,n}(q,p)$ is the amplitude, must be introduced for this reason.

Notice that the probability defines the time reversible motion. Therefore, (i) to have a possibility to describe correctly the quantum interference effects and, at the same time, (ii) to have the quantum processes time reversible, one must introduce the amplitudes and must suppose that their norm are measurable quantities only.

One must take into account the conservation laws. It is useful to introduce the factor

$$\prod_{j=1}^{m} z^i(q_j) \prod_{j=1}^{n} z^f(p_j),$$

to have a possibility to regulate the number of in- and outgoing particles. We distinguish the initial and final states and for this reason we will have two Lagrange multipliers of the energy-momentum conservation laws. For this reason we introduce

$$\prod_{j=1}^{m} e^{-i\alpha^i q_j} \prod_{j=1}^{n} e^{-i\alpha^f p_j}.$$

The two factors depending on z and α are the *probe functions* through which we will extract Ω, i.e. model the measuring devices.

As a result,

$$d\Gamma_{mn} = d\omega_m(q;\alpha^i,z^i)d\omega_n(p;\alpha^f,z^f)|a_{m,n}(q,p)|^2, \tag{4}$$

where

$$d\omega_m(q;\alpha^i,z^i) = \prod_{j=1}^{m}\frac{dq_j z^i(q_j)e^{-i\alpha^i q_j}}{(2\pi)^3 2\varepsilon_j\, m!},\ \ \varepsilon_j = +\sqrt{q_j^2+m^2} \tag{5}$$

and $d\omega_n$ has the same definition.

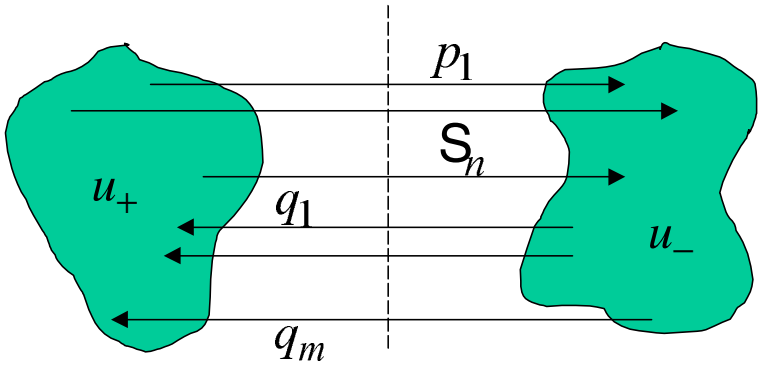


Fig. 6: The diagram for $\rho(\alpha,z)$. Each incoming line carries the factor $z_i(q_k)e^{-i\alpha_i q_k}$, $k=1,2,...,m$, and the outgoing one carries $z_i(p_k)e^{-i\alpha_i p_k}$, $k=1,2,...,n$. The field u_+ is defined on the Mills complex time contour C and u_- is defined on C^*. The summation over the number of particles, n and m, is assumed

The definition of ρ is given in Eq. (2). The rule for ρ is constructed shown in Fig. 6. The calculations are described in Appendix A and the result of integrations is as follows:

$$\rho(\alpha,z) = \exp\{i\int dxdx'(\hat{u}_+(x)D_{+-}(x-x',\alpha^f,z^f)\hat{u}_-(x')-$$

$$-\hat{u}_-(x)D_{-+}(x-x',\alpha^i,z^i)\hat{u}_+(x'))\}\,\rho_0(u),\ \ \hat{u}(x)\equiv\frac{\delta}{\delta u(x)},$$

where

$$\rho_0(\phi) = Z(u_+)Z^*(u_-)$$

and D_{ij} are the frequency Green functions if z is equal to 1, see Eqs. (A.8), (A.9). Calculating ρ_0 perturbatively, one can find:

$$\rho(\alpha,z) = \exp\{-iV(-i\hat{j}_+)+iV(-i\hat{j}_-)\}\times$$

$$\times\exp\{\frac{i}{2}\int dxdx' j_i(x)G_{ij}(x-x',\alpha,z)j_j(x')\},\ \ \hat{j}(x)\equiv\frac{\delta}{\delta j(x)} \tag{6}$$

and the summation over the repeated indexes is assumed, D_{++} is the Feynman causal Green function and $D_{--}=D_{++}^*$ is the anticausal Green function.

Taking $t_i=-\infty$ and $t_f=+\infty$ and calculating the integral (8) perturbatively we find the coincidence of ρ_{cr} defined in Eq. (A.27) and ρ from Eq. (6) $if\ \beta_i=\beta_f$. The distinction is hidden in the definition of the Green functions.

The 'factorization' of contributions from contours C_+ and C_- in the integral (8) follows from the Rieman–Lebesgue lemma [14] which is applicable in the perturbation framework [15]. Note the absence of the Matsubara parts of the contour on the remote hypersurface, $|t|\to\infty$.

2.2 Schwinger–Keldysh formalism

There are various approaches to building the real-time finite-temperature field theories of the Schwinger–Keldysh type (see, for example Ref. [14]). All of them use various tricks for analytical continuation of imaginary-time the Matsubara formalism to the real time. The basis of the approaches is the introduction of the Matsubara field operator

$$\Phi_M(\mathbf{x},\beta) = e^{\beta H}\Phi_S(\mathbf{x})e^{-\beta H}, \tag{7}$$

where $\Phi_S(\mathbf{x})$ is the interaction-picture operator, instead of the Heisenberg operator

$$\Phi(\mathbf{x},t) = e^{itH}\Phi_S(\mathbf{x})e^{-itH}.$$

This introduces the averaging over the Gibbs ensemble instead of averaging over zero-temperature vacuum states.

If the interaction switched on adiabatically at the instant t_i and switched off at t_f then there is the unitary transformation:

$$\Phi(x) = U(t_i,t_f)U(t_i,t)\Phi_S(x)U(t,t_i).$$

Introducing the complex Mills time contours [16] to connect t_i to t, t to t_f and t_f to t_i we form the 'closed-time' contour C (the end-points of the contours C_+ and C_- are $joined$ together), see Fig. 7. This allows the last equality to be written in the compact form:

$$\Phi(x) = T_C\{\Phi(x)e^{i\int_C d^4x' L_{int}(x')}\}_S,$$

where T_C is the time-ordering on the contour C operator.

The corresponding expression for the generating functional $Z(j)$ of Green functions has the form:

$$Z(j) = R(0) < T_C e^{i\int_C d^4x\{L_{int}(x)+j(x)\Phi(x)\}_S} >,$$

where $<>$ means averaging over the initial state.

If the initial correlations have a little effect we can perform averaging over the Gibbs ensemble. This is the main assumption of formalism: the generating functional of the Green functions $Z(j)$ has the form:

$$Z(j) = \int D\Phi' < \Phi';t_i|e^{-\beta H}T_C e^{i\int_C d^4x j(x)\Phi(x)}|\Phi';t_i >$$

with $\Phi' = \Phi'(\mathbf{x})$. In accordance with Eq. (7) we have:

$$< \Phi';t_i|e^{-\beta H} =< \Phi';t_i - i\beta|$$

and, as a result,

$$Z(j) = \int D\Phi' e^{i\int_{C_\beta} d^4x\{L(x)+j(x)\Phi(x)} \tag{8}$$

where the path integration is performed with the KMS periodic boundary condition:

$$\Phi(t_i) = \Phi(t_i - i\beta).$$

In Eq. (8) the contour C_β connects t_i to t_f, t_f to t_i and t_i to $t_i - i\beta$. Therefore it contains the imaginary-time Matsubara part t_i to $t_i - i\beta$. A more symmetrical formulation uses the following realization: t_i to t_f, t_f to $t_f - i\beta/2$, $t_f - i\beta/2$ to $t_i - i\beta/2$ and $t_i - i\beta/2$ to $t_i - i\beta$ (e.g. [15]), see Fig. 5. This case also contains the imaginary-time parts of the time contour. Therefore, Eq. (8) presents the analytical continuation of the Matsubara generating functional to real times.

One can note that if this analytical continuation is possible in $Z(j)$ then the representation (8) gives a good recipe for regularization of frequency integrals in the Matsubara perturbation theory, see, for example, Ref. [14], but there is nothing new for our problem since the Matsubara formalism is a formalism for equilibrium states only.

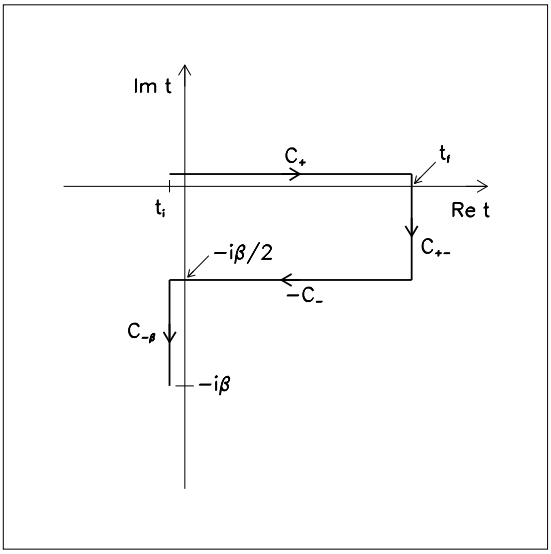


Fig. 7: Niemi–Semenoff time contour.

2.3 S-matrix: relaxation of correlations

Let us return now to Eq. (4). It is useful to take $\alpha = (-i\beta, \vec{0})$. This is possible if the number of particles is sufficiently large [17].

The value of β can be defined through the equations of state (A.14). In this case we will consider $1/\beta_{i(f)}$ as the mean energy of particles in the initial (final) state. But even knowing solutions of this equations one can not find $\rho(E, z)$ correctly if the fluctuations near the solutions of Eq.(A.14) are not Gaussian, i.e. if the energy is not distributed uniformly over the degrees of freedom.

Let us define the conditions when the fluctuations in the vicinity of β_c are small. Firstly, to estimate the integral (A.5) in the vicinity of the extremum β_c we should expand $\ln \rho_n(\beta + \beta_c)$ over β:

$$\ln \rho_n(\beta + \beta_c) = \ln \rho_n(\beta_c) - \sqrt{s}\beta + \frac{1}{2!}\beta^2 \frac{\partial^2}{\partial \beta_c^2} \ln \rho_n(\beta_c) - \frac{1}{3!}\beta^3 \frac{\partial^3}{\partial \beta_c^3} \ln \rho_n(\beta_c) + ... \tag{9}$$

and, secondly, expand the exponent over, for instance, $\partial^3 \ln \rho_n(\beta_c)/\partial \beta_c^3$, if higher terms in Eq. (9) are neglected. As a result, the k-th term of the perturbation series

$$\rho_{n,k} \sim \left\{ \frac{\partial^3 \ln \rho_n(\beta_c)/\partial \beta_c^3}{(\partial^2 \ln \rho_n(\beta_c)/\partial \beta_c^2)^{3/2}} \right\}^k \Gamma\left(\frac{3k+1}{2}\right). \tag{10}$$

Therefore, the perturbation theory near β_c leads to the asymptotic series. The formal assumption that one can define this series, for instance, in the Borel sense is not interesting from the physical point of view. Indeed, this solution assumes that the fluctuations near β_c are arbitrarily high and as a result the quantity β_c loses its significance.

But we may assume that

$$\partial^3 \ln \rho_n(\beta_c)/\partial \beta_c^3 << (\partial^2 \ln \rho_n(\beta_c)/\partial \beta_c^2)^{3/2}, \tag{11}$$

to neglect this term. One possible solution of this condition is

$$\partial^3 \ln \rho_n(\beta_c)/\partial \beta_c^3 \approx 0. \tag{12}$$

If this condition holds, then the fluctuations are Gaussian with dispersion

$$\sim \{\partial^2 \ln \rho_n(\beta_c)/\partial\beta_c^2\}^{1/2},$$

see Eq. (9).

Let us consider now (12) carefully. We will find, computing derivatives, that this condition means the following approximate equality:

$$\frac{\rho_n^{(3)}}{\rho_n} - 3\frac{\rho_n^{(2)}\rho_n^{(1)}}{\rho_n^2} + 2\frac{(\rho_n^{(1)})^3}{\rho_n^3} \approx 0, \tag{13}$$

where $\rho_n^{(k)}$ means the k-th derivative. For identical particles,

$$\rho_n^{(k)}(\beta_c) = n^k(-1)^k \int \left\{\prod_{i=1}^{n} \frac{d^3q_i e^{-\beta\epsilon(q_i)}}{(2\pi)^3 2\epsilon(q_i)}\right\} |a_n|^2 \prod_{i=1}^{k} \epsilon(q_i) =$$

$$= \int \left\{\prod_{i=1}^{k} \epsilon(q_i) \frac{d^3q_i e^{-\beta\epsilon(q_i)}}{(2\pi)^3 2\epsilon(q_i)}\right\} \bar{f}_n^{(k)}(q_1, q_2, ..., q_k), \tag{14}$$

where $\bar{f}_n^{(k)}$ coincides with the k-particle momentum distribution function if the n-particle system is produced. Therefore, the l.h.s. of (13) is the 3-point energy correlator K_3:

$$K_3 \equiv \int d\omega_3(q) \left(< \prod_{i=1}^{3} \epsilon(q_i) >_{\beta_c} -3 < \prod_{i=1}^{2} \epsilon(q_i) >_{\beta_c} < \epsilon(q_3) >_{\beta_c} +2 \prod_{i=1}^{3} < \epsilon(q_i) >_{\beta_c} \right), \tag{15}$$

where the index means averaging with the Boltzmann factor $\exp\{-\beta_c\epsilon(q)\}$.

As a result, to have all fluctuations in the vicinity of β_c Gaussian, we should have $K_l \approx 0$, $l \geq 3$. Notice that, as follows from (11), the set of minimal conditions actually looks as follows:

$$|K_l|^{2/l} << K_2,\ l \geq 3. \tag{16}$$

If the experiment confirms these conditions then, independently from the number of produced particles,

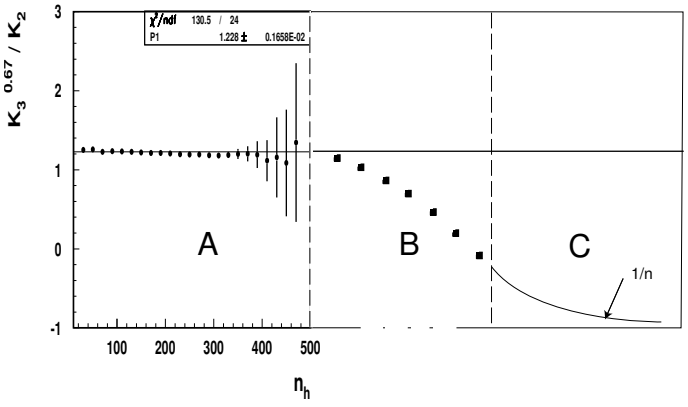


Fig. 8: PYTHIA prediction.

the final state may be described with high enough accuracy by one parameter β_c. The Monte Carlo

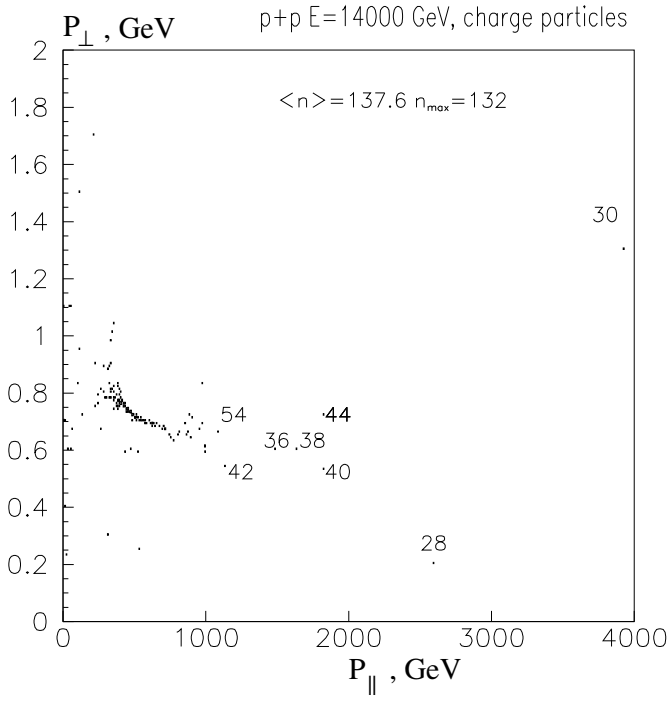


Fig. 9: PYTHIA prediction for ration of the mean values, $p_\perp/p_{||}$. This ratio is equal to $4/\pi$ for equilibrium kinematics. The numbers are the corresponding multiplicities.

prediction for the ratio $R = |K_3|^{2/3}/|K_2|$ based on PYTHIA is shown in Fig. 8. Figure 9 shows the PYTHIA prediction for the $(p_\perp/, p_{||})$ plot. Both 'experiments' show the absence of thermalization in the accessible domain of multiplicities.

The condition (16) is a formal consequence of zero convergence radii of the perturbation series near β_c. Therefore, the condition (16) is unique if the notion of the temperature $1/\beta_c$ exists.

Considering β_c as a physical (measurable) quantity, we are forced to assume that both the total energy of the system $\sqrt{s} = E$ and conjugate to it the variable β_c may be measured with high accuracy[1].

2.4 Temperature fluctuations in the S-matrix approach

It is not necessary to measure the energy of each particle to have $n > \bar{n}$. Indeed, let $\tilde{\epsilon}_i$ be the energy of the i-th cell of particles, $\tilde{\epsilon}_1 + \tilde{\epsilon}_2 + ... + \tilde{\epsilon}_k = \sqrt{s}$ and let $\tilde{n}_i$ be the number of particles in the cell, $\tilde{n}_1 + \tilde{n}_2 + ... + \tilde{n}_k = n$. Then, if $\tilde{\epsilon}_i < \epsilon_0$, $i = 1, 2, ..., k$, we have the inequality: $k > n_{min}$. Therefore, we get into the high multiplicities domain since $n \geq k$, if $\epsilon_0 < \sqrt{s}/\bar{n}(s)$. We can use the calorimeter provided that the energy $\tilde{\epsilon}_i < \epsilon_0$ in each cell.

The preparation of such an experiment is not a hopeless task and it may be sufficiently informative. We will put this formulation of the experiment in the basis of the theory. These forces use the Wigner functions formalism and the first question that must be solved is how to find a way to adopt this formalism for the description of our experiment.

We start the consideration from the assumption that the temperature fluctuations are large scale. In a cell whose dimension is much smaller than the fluctuation scale of the temperature we can assume that the temperature is a 'good' parameter. (The 'good' parameter means that the corresponding fluctuations are Gaussian.)

[1]Note, the uncertainty principle did not restrict ΔE and $\Delta\beta$.

Let us surround the interaction region, i.e. the system under consideration, with N cells with known space-time position and let us propose that we can measure the energy and momentum of groups of incoming and outgoing particles in each cell. The fourth-dimension of the cells can not be arbitrarily small in this case because of the quantum uncertainty principle.

Calculating ρ_0 perturbatively we will find in the case of a periodic boundary condition that

$$\rho_{cp}(\beta) = \exp\{-iV(-i\hat{j}_+) + iV(-i\hat{j}_-)\} \times$$
$$\exp\{i\int drdy[\hat{j}_i(r+y/2)G_{ij}(y,(\beta(r))\hat{j}_j(r-y/2)\} \tag{17}$$

where, using the matrix notations,

$$iG(q,(\beta(r))) = \begin{pmatrix} \frac{i}{q^2-m^2+i\varepsilon} & 0 \\ 0 & -\frac{i}{q^2-m^2-i\varepsilon} \end{pmatrix} +$$
$$+2\pi\delta(q^2-m^2)\begin{pmatrix} n(\frac{(\beta_1+\beta_2)(r)}{2}|q_0|) & n(\beta_1(r)|q_0|)a_+(\beta_1) \\ n(\beta_2(r)|q_0|)a_-(\beta_2) & n(\frac{(\beta_1+\beta_2)(r)}{2}|q_0|) \end{pmatrix}, \tag{18}$$

and

$$a_\pm(\beta) = -e^{\beta(|q_0|\pm q_0)/2},$$

the occupation number

$$n(\beta q_0) = \frac{1}{e^{\beta q_0}-1}.$$

The details can be found in Appendix B.

It is shown in Appendix C that r in Eq. (17) is the measurement point of the temperature. It is assumed that the fluctuations are Gaussian in each point r.

3 CONCLUSIONS

Introducing the temperature as the Lagrange multiplier, we should assume that the temperature fluctuations are small (Gaussian). In the opposite case the notion of temperature loses its sense. The 'working' idea about nonequilibrium processes is based on the assumption that the evolution of a system goes through a few phases. In the first 'fast' phase the s-particle distribution functions $\mathbf{D}_s$, $s>1$, strongly depends on the initial conditions. But at the end of this phase the system forgets the initial-state information. The second phase is the 'kinetic' one. One can expect that the space–time fluctuations of thermodynamical parameters in this phase are large scale, i.e. there are macroscopic domains in which the subsystems are in equilibrium, with Gaussian fluctuations of the thermodynamical parameters. In the last "hydrodynamical' phase the whole system is described by macroscopic parameters. We shall see that the Schwinger–Keldysh [14, 18] formalism is applicable for the 'hydrodynamical' phase only.

The S-matrix finite-temperature description can be realized not only for the uniform temperature distribution (we have taken the first step in this direction wishing to introduce initial and final temperatures separately). So, by introducing cells of a measuring device (calorimeter) and introducing the energy-momentum shells of each cell separately, we can introduce the individual temperatures in each cell. This can be done since in the S-matrix theory the measurement is performed by free (mass-shell) particles, i.e. the measurement of energy (and momentum) can be performed in each cell separately. This allows one to capture the 'kinetic' phase also (if the number of calorimeter cells is high enough). In this phase multiparticle distribution functions $\mathbf{D}_s$, $s>1$, are functionals of the one-particle distribution function $\mathbf{D}_1$ only. This means the 'shortened' description of the nonequilibrium medium [11].

The logic of our construction of the thermodynamics is the following.

A. Let us assume that, for definiteness, a given system can be characterized by the 'good' parameter β_c which is conjugate to the energy of a produced particle, i.e. let us assume that the fluctuations in the vicinity of β_c are Gaussian. In this case we can construct a closed perturbation theory for the generating functional of observables $\rho(\beta_c, z)$. The parameter z will be defined later.

Then, if the system is surrounded by black-body radiation, $\rho(\beta_c, z)$ coincides with the generating functional of the correlation function in the real-time finite-temperature statistics of Schwinger–Keldysh [18]. We can also apply the vacuum boundary condition. The corresponding formalism can be applied to particle physics.

The following step consists in the assumption that there are no correlations on the remote hypersurface σ_∞. Then one may analytically continue $\rho(\beta_c, z)$ to the generating functional in the Matsubara formalism [19], i.e. $\rho(\beta_c, z)$ would coincide with the partition function, where β_c is the inverse temperature, $T_c = 1/\beta_c$, and $T_c \ln z = \mu$ is the chemical potential, i.e. z is the activity. But at this stage we can not use the vacuum boundary conditions.

Therefore, it is impossible to introduce the canonical formalism for the description of accelerator experiments because of the absence of the thermostat ('heat bath'). Nevertheless the thermodynamical method, based on the introduction of 'good rough' variables, β_c, μ, etc., can be applied.

B. In the canonical formalism the temperature is introduced using the periodic boundary condition of Kubo, Martin and Schwinger (KMS) for a field [10]:

$$\Phi(t) = \Phi(t - i\beta).$$

But, without fail, this method leads to *equilibrium* fluctuation-dissipation conditions [20] (see also Ref. [21]). Moreover, it leads to the quantization of the energy spectrum and additional troubles in transitions to the continuous limit, see, for example, Ref. [14]. This is why the real-time formalism is useful. In our case the KMS boundary condition would be the consequence of Gaussian fluctuations in the vicinity of β_c and of the assumption that the system is surrounded by black body radiation.

Notice also that it is not clear whether the KMS boundary condition can be applied in the case of local temperature $\beta_c = \beta_c(x, t)$. Having no necessity to use the KMS boundary condition, we will extend the formalism to the case of $\beta_c = \beta_c(x, t)$. This will allow us to describe the kinetic stage of the dissipation process [22].

C. We can consider the equilibrium over the other parameters. For instance, the equilibrium over the charge of produced particles. So, if an initial state is neutral, i.e. if, for example, the Tevatron experiments are considered, then the charge conservation means that the algebraic sum of a number of the positive charged particles, n_+, and negative charged particles, n_-, must be equal to zero, $n_+ - n_- = 0$. At the same time the value of n_+, n_- and a number of neutral particles, n_0, can be arbitrary, i.e. these quantities are fluctuating parameters.

Instead of the temperature T it is useful to introduce the quantity $\beta = 1/T$ which is the Lagrange multiplier of the energy conservation law. Accordingly, one may introduce $\mu_i, \;\; i = +, -, 0$, which has the notion of chemical potential and is conjugate to the corresponding numbers $n_i, \;\; i = +, -, 0$. The equilibrium over the charge means that $\mu_+ = \mu_- = \mu_0$, i.e. $n_i = n/3$, where n is $total$ multiplicity. In the case of equilibrium the fluctuations $must$ be Gaussian. The reason, why the fluctuations of numbers n_i are not Gaussian was offered in Ref. [23]. Instead of the examination of the Gaussian distribution one can examine the charge correlators.

APPENDIX A: FINITE-TEMPERATURE S-MATRIX FORMALISM

A1 *VACUUM BOUNDARY CONDITIONS*

The starting point of our calculations is the n- into m-particles transition amplitude $a_{n,m}$, the derivation of which is a well-known procedure in the perturbation theory framework. For this purpose the $(n+m)$-point Green function $G_{n,m}$ is introduced. To calculate the nontrivial elements of the S-matrix one must put the external particles on the mass shell. Formally, this procedure means amputation of the external legs of $G^c_{n,m}$. As a result the amplitude of the m- into n-particles transition $a_{n,m}$ in the momentum representation has the form:

$$a_{n,m}((q)_n;(p)_m) = (-i)^{n+m}\prod_{k=1}^{m}\hat{u}(q_k)\prod_{k=1}^{n}\hat{u}^*(p_k)Z(u). \tag{A.1}$$

Here we introduce the 'annihilation' operator

$$\hat{u}(q) = \int dx e^{-iqx}\hat{u}(x), \quad \hat{u}(x) = \frac{\delta}{\delta u(x)}, \tag{A.2}$$

$\hat{u}^*(p_k)$ is the 'creation' operator and q_k and p_k are the momentum of the incoming and outgoing particles. In (A.1)

$$Z(u) = \int D\Phi e^{iS(\Phi)-iV(\Phi+u)}$$

is the generating functional. The total action was divided into two parts, where $S(\Phi)$ is a free part and $V(\Phi,\phi)$ describes the interactions. At the very end one should put the auxiliary field $\phi = 0$.

To provide the convergence of the integral (A.1) over the scalar field Φ, the action $S(\Phi)$ must contain a positive imaginary part. Usually for this purpose Feynman's $i\varepsilon$-prescription is used. It is better for us to shift the time contour infinitesimally to the upper half plane [14, 16], i.e. to the Mills contour

$$C_+ : t \to t + i\varepsilon, \quad \varepsilon > 0$$

and after all calculations to return the time contour on the real axis, $\varepsilon \to +0$.

In Eq. (A.1) the integration is performed over all field configurations with a standard vacuum boundary condition:

$$\int d^4x \partial_\mu(\Phi\partial^\mu\Phi) = \int_{\sigma_\infty} d\sigma_\mu \Phi\partial^\mu\Phi = 0,$$

which assumes zero contribution from the surface term.

Supposing that the particle number and momenta are insufficient for us we introduce the probability

$$r(P) = \sum_{n,m}\frac{1}{n!m!}\int d\omega_n(q)d\omega_m(p)\delta^{(4)}(P-\sum_{k=1}^{n}q_k)\delta^{(4)}(P-\sum_{k=1}^{n}p_k)|a_{n,m}|^2, \tag{A.3}$$

where

$$d\omega_n(q) = \prod_{k=1}^{n}d\omega(q_k) = \prod_{k=1}^{n}\frac{d^3q_k}{(2\pi)^3 2\epsilon(q_k)}, \quad \epsilon = (q^2+m_h^2)^{1/2},$$

is the Lorentz-invariant phase space element. We assume that the energy–momentum conservation δ-function was extracted from the amplitude. It was divided into two parts:

$$\delta^{(4)}(\sum q_k - \sum p_k) = \int d^4P\delta^{(4)}(P-\sum q_k)\delta^{(4)}(P-\sum p_k). \tag{A.4}$$

It is not too hard to see that, up to the phase space volume,

$$r = \int d^4P r(P)$$

is the imaginary part of the amplitude $< vac|vac >$. Therefore, by computing $r(P)$ the standard renormalization procedure can be applied and the new divergences will not arise in our formalism.

The Fourier transformation of δ-functions in (A.3) allows us to write $r(P)$ in the form:

$$r(P) = \int \frac{d^4\alpha_1}{(2\pi)^4}\frac{d^4\alpha_2}{(2\pi)^4} e^{iP(\alpha_1+\alpha_2)}\rho(\alpha_1,\alpha_2), \tag{A.5}$$

where

$$\rho(\alpha_1,\alpha_2) = \sum_{n,m}\frac{1}{n!m!}\int\prod_{k=1}^{n}\{d\omega(q_k)e^{-i\alpha_1 q_k}\}\prod_{k=1}^{m}\{d\omega(p_k)e^{-i\alpha_2 p_k}\}|a_{n,m}|^2. \tag{A.6}$$

The introduction of the 'Fourier-transformed' probability $\rho(\alpha_1,\alpha_2)$ means only that the phase-space volume is not fixed exactly, i.e. it is proposed that the 4-vector P is fixed with some accuracy if α_i is fixed. The energy and momentum in our approach are still locally conserved quantities since the amplitude a_{nm} is translational invariant. So, we can perform the transformation:

$$\alpha_1\sum q_k = (\alpha_1-\sigma_1)\sum q_k + \sigma_1\sum q_k \to (\alpha_1-\sigma_1)\sum q_k + \sigma_1 P$$

since four-momenta are conserved. The choice of σ_1 fixes the reference frame. This degree of freedom of the theory was considered in Ref. [24].

Inserting (A.1) into (A.6) we find that

$$\begin{aligned}\rho(\alpha_1,\alpha_2) = \exp\{i\int dxdx'(\hat{u}_+(x)D_{+-}(x-x',\alpha_2)\hat{u}_-(x') - \\ -\hat{u}_-(x)D_{-+}(x-x',\alpha_1)\hat{u}_+(x'))\}Z(u_+)Z^*(u_-),\end{aligned} \tag{A.7}$$

where D_{+-} and D_{-+} are the positive and negative frequency correlation functions:

$$D_{+-}(x-x',\alpha) = -i\int d\omega(q)e^{iq(x-x'-\alpha)} \tag{A.8}$$

describes the process of particle creation at the time moment x_0 and its absorption at x_0', $x_0 > x_0'$, and α is the centre of mass () 4-coordinate. The function

$$D_{-+}(x-x',\alpha) = i\int d\omega(q)e^{-iq(x-x'+\alpha)} \tag{A.9}$$

describes the opposite process, $x_0 < x_0'$. These functions obey the homogeneous equations:

$$(\partial^2+m^2)_x G_{+-} = (\partial^2+m^2)_x G_{-+} = 0$$

since the propagation of mass-shell particles is described.

We suppose that $Z(\phi)$ may be computed perturbatively. For this purpose the following transformation will be used:

$$\begin{aligned}e^{-iV(u)} = e^{-i\int dx\hat{j}(x)\hat{u}'(x)}e^{i\int dxj(x)u(x)}e^{-iV(u')} = \\ = e^{\int dxu(x)\hat{u}'(x)}e^{-iV(u')} = \\ = e^{-iV(-i\hat{j})}e^{i\int dxj(x)u(x)},\end{aligned} \tag{A.10}$$

where $\hat{u}$ was defined in (A.2). At the end of the calculations the auxiliary variables j, ϕ' should be taken equal to zero. Using the first equality in (A.10) we find that

$$Z(u) = e^{-i\int dx\hat{j}(x)\hat{u}(x)}e^{-iV(u+\phi)}e^{-\frac{i}{2}\int dxdx'j(x)D_{++}(x-x')j(x')}, \tag{A.11}$$

where D_{++} is the causal Green function:

$$(\partial^2 + m^2)_x G_{++}(x-y) = \delta(x-y).$$

Inserting (A.11) into (A.7) after simple manipulations with differential operators, see (A.10) we find the expression:

$$\begin{aligned}\rho(\alpha_1,\alpha_2) = e^{-iV(-i\hat{j}_+)+iV(-i\hat{j}_-)} \times \\ \times \exp\{\frac{i}{2}\int dxdx'(j_+(x)D_{+-}(x-x',\alpha_1)j_-(x') - \\ j_-(x)D_{-+}(x-x',\alpha_2)j_+(x') - \\ -j_+(x)D_{++}(x-x')j_+(x') + j_-(x)D_{--}(x-x')j_-(x'))\},\end{aligned} \tag{A.12}$$

where

$$D_{--} = (D_{++})^*$$

is the anticausal Green function.

Considering the system with a large number of particles we can simplify calculations choosing the CM frame $P = (P_0 = E, \vec{0})$. It is useful also [17] to rotate the contours of integration over $\alpha_{0,k}$: $\alpha_{0,k} = -i\beta_k, Im\beta_k = 0, k = 1,2$. As a result, omitting the unnecessary constant, we will consider $\rho = \rho(\beta_i, \beta_f)$. The external particles play a double role in the S-matrix approach: their interactions create and annihilate the interacting fields system and, on the other hand, they are probes through which the measurement of the system is performed. Since β_k are the conjugate to the quantities of particle energy we will interpret them as the inverse temperatures in the initial (β_+) and final (β_-) states of interacting fields. But there is the question: Are constants β_k really the 'good' parameters to describe the system?

The integrals over β_k:

$$r(E) = \int \frac{d\beta_1}{2\pi i}\frac{d\beta_2}{2\pi i} e^{(\beta_1+\beta_2)E} e^{-F(\beta_1,\beta_2)}, \tag{A.13}$$

where

$$F(\beta_1,\beta_2) = -\ln\rho(\beta_1,\beta_2),$$

can be computed by the stationary phase method. This assumes that the total energy E is a fixed quantity. The solutions of the equations (of state):

$$E = \frac{\partial F(\beta_1,\beta_2)}{\partial\beta_k}, \quad k = 1,2, \tag{A.14}$$

gives the most probable values of β_k at a given E. Equations (A.14) always have the real solutions and, because of the energy conservation law, both Eqs. (A.14) have the same solution with the property:

$$\beta_k = \beta(E), \quad \beta > 0.$$

Assuming that β is the 'good' parameter, i.e. the fluctuations of β_k are Gaussian, we can interpret $F(\beta_i,\beta_f)$ as the free energy and $1/\beta_k$ as the temperatures. Such a definition of thermodynamical parameters is in the spirit of microcanonical description.

The structure of the generating functional (A.12) is the same as the generating functional of Niemi–Semenoff [15]. The difference is only in the definition of the Green functions which follows from the choice of the boundary condition. The Green functions $D_{ij}, i,j = +,-$ were defined on the time contours $C_\pm$ in the complex time plane ($C_- = C_+^*$). This definition of the time contours coincides with Keldysh's time contour [18]. The expression (A.12) can be written in the compact form if the matrix notations are used. Note also a doubling of the degrees of freedom. This doubling is unavoidable since the Green functions D_{ij} are singular on the light cone.

A2 *CLOSED-PATH BOUNDARY CONDITIONS*

The generating functional $\rho(\alpha_1, \alpha_2)$ has an important factorized structure, see (A.7):

$$\rho(\alpha_1, \alpha_2) = e^{\hat{N}(\alpha_1, \alpha_2; \phi)} \rho_0(\phi_\pm),$$

where the operator

$$\hat{N}(\alpha_1, \alpha_2; \phi) = \int dx dx' (\hat{\phi}_+(x) D_{+-}(x - x', \alpha_2) \hat{\phi}_-(x') - \\ - \hat{\phi}_-(x) D_{-+}(x - x', \alpha_1) \hat{\phi}_+(x')) \tag{A.15}$$

acts on the generating functional

$$\rho_0(\phi_\pm) = Z(\phi_+) Z^*(\phi_-) = \\ = \int D\Phi_+ D\Phi_- e^{iS(\Phi_+) - iS(\Phi_-) - iV(\Phi_+ + \phi_+) + iV(\Phi_- + \phi_-)}, \tag{A.16}$$

of measurables. All 'thermodynamical' information is contained in the operator $\hat{N}(\alpha_1, \alpha_2; \phi)$ and interactions are hidden in $\rho_0(\phi_\pm)$. One may say that the action of the operator $\hat{N}$ maps the system of interacting fields on the measurable states. The last ones are 'labelled' by α_1 and α_2. Just this property allows us to say that we are dealing with 'mechanical' fluctuations only. To regulate the number of particles we can introduce into $\hat{N}$ the dependence from 'activities' z_1 and z_2 for initial and final states separately.

The independent fields ϕ_+, ϕ_- and Φ_+, Φ_- were defined on the time contours C_+, C_-. By definition, the path integral (A.16) describes the closed path motion in the space of fields Φ. We want to use this fact and introduce a more general boundary condition which also guarantees cancellation of surface terms in the perturbation framework. We will introduce the equality:

$$\int_{\sigma_\infty} d\sigma_\mu \Phi_+ \partial^\mu \Phi_+ = \int_{\sigma_\infty} d\sigma_\mu \Phi_- \partial^\mu \Phi_-. \tag{A.17}$$

The solution of Eq.(A.17) requires that the fields Φ_+ and Φ_- (and theirs first derivatives $\partial_\mu \Phi_\pm$) coincide on the boundary hypersurface σ_∞:

$$\Phi_\pm(\sigma_\infty) = \Phi(\sigma_\infty),$$

where, by definition, $\Phi(\sigma_\infty)$ is the arbitrary, 'turning-point', field.

The existence of the nontrivial field $\Phi(\sigma_\infty)$, in the absence of surface terms, has an influence only on the structure of Green functions

$$G_{++} = < T\Phi_+\Phi_+ >, \quad G_{+-} = < \Phi_+\Phi_- >,$$

$$G_{-+} = < \Phi_-\Phi_+ >, \quad G_{--} = < \tilde{T}\Phi_-\Phi_- >,$$

where $\tilde{T}$ is the antitemporal time ordering operator. These Green functions must obey the equations:

$$(\partial^2 + m^2)_x G_{+-}(x - y) = (\partial^2 + m^2)_x G_{-+}(x - y) = 0,$$

$$(\partial^2 + m^2)_x G_{++}(x - y) = (\partial^2 + m^2)^*_x G_{--}(x - y) = \delta(x - y),$$

and the general solution of these equations:

$$G_{ii} = D_{ii} + g_{ii}, \\ G_{ij} = g_{ij}, \quad i \neq j \tag{A.18}$$

contains the undefined terms g_{ij} which must obey the homogenous equations:

$$(\partial^2 + m^2)_x g_{ij}(x-y) = 0, \quad i,j = +,-. \tag{A.19}$$

The general solution of these equations (they are distinguished by the choice of the time contours $C_\pm$)

$$g_{ij}(x-x') = \int d\omega(q) e^{iq(x-x')} n_{ij}(q) \tag{A.20}$$

are defined by the functions n_{ij}. The last ones are the functionals of the 'turning-point' field $\Phi(\sigma_\infty)$: if $\Phi(\sigma_\infty) = 0$ we must have $n_{ij} = 0$ and we will come back to the theory of the previous section.

Our aim is to define n_{ij}. We can suppose that

$$n_{ij} \sim < \Phi(\sigma_\infty) \cdots \Phi(\sigma_\infty) > .$$

The simplest supposition gives:

$$n_{ij} \sim < \Phi_i \Phi_j > \sim < \Phi^2(\sigma_\infty) > . \tag{A.21}$$

We will find the exact definition of n_{ij} starting from the S-matrix interpretation of the theory.

We should suppose that there are only free, mass-shell, particles on the infinitely far hypersurface σ_∞. Formally this follows from (A.18)–(A.20) and is natural in the S-matrix framework [13]. In other respects the choice of the boundary condition is arbitrary.

Therefore, our aim is the description of evolution of the system in a background field of mass-shell particles. We will assume that there are not any special correlations among background particles and we will take into account only the constraints of the energy-momentum conservation laws. Quantitatively this means that the multiplicity distribution of the background particles is Poisson-like, i.e. is determined by the mean multiplicity only. This is in the spirit of the definition of n_{ij} in Eqs. (A.20), (A.21).

Our derivation is the same as in Ref. [24]. Here we restrict ourselves mentioning only the main quantitative points.

In the vacuum case the process of particle creation and their further absorption was described. In the presence of the background particles this time-ordered picture is wiped out: there appears a possibility of particle absorption before their creation.

The particle creation and absorption was described by the product of an operator exponent (A.7). One can derive (see also Ref. [24]) the generalizations of (A.7): the presence of the background particles will lead to the same structure:

$$\rho_{cp} = e^{i\hat{N}(\phi_i^* \phi_j)} \rho_0(\phi_\pm),$$

where $R_0(\phi_\pm)$ is the same generating functional, see (A.16). But the operator $\hat{N}(\phi_i^* \phi_j)$, $i,j = +,-$, should be changed to take into account the external particle environment.

The operator $\hat{\phi}_i^*(q)$ was interpreted as the creation and $\hat{\phi}_i(q)$ as the annihilation operator, see definition (A.1). Correspondingly the product $\hat{\phi}_i^*(q)\hat{\phi}_j(q)$ acts as the activity operator. So, in the expansion of $\hat{N}(\phi_i^* \phi_j)$ we can leave only the first nontrivial term:

$$\hat{N}(\phi_i^* \phi_j) = \int d\omega(q) \hat{\phi}_i^*(q) n_{ij} \hat{\phi}_j(q), \tag{A.22}$$

since no special correlation among background particles should be expected. If the external (nondynamical) correlations are present then the higher powers of $\hat{\phi}_i^* \hat{\phi}_j$ will appear in the expansion (A.22) [25]. Following the interpretation of $\hat{\phi}_i^* \hat{\phi}_j$ we conclude that n_{ij} is the mean multiplicity of the background particles. In (A.22) the normalization condition $N(0) = 0$ was used and the summation over all i,j was assumed. (In the vacuum case only the combinations $i \neq j$ was present.)

Computing R_{cp} we must conserve the translational invariance of amplitudes and extract the energy-momentum conservation δ-functions. We must adjust to each vertex of an incoming particle in $a_{n,m}$ the factor $e^{-i\alpha_1 q/2}$ and for each outgoing particle $e^{-i\alpha_2 q/2}$ one post Fourier transformation of the δ-functions.

So, the product $e^{-i\alpha_k q/2}e^{-i\alpha_j q/2}$ can be interpreted as the probability factor of the one-particle $(creation + annihilation)$ process. The n-particle $(creation + annihilation)$ process probability is the simple product of these factors if there are not special correlations among background particles. This interpretation is evident in the CM frame $\alpha_k = (-i\beta_k, \vec{0})$.

After these preliminaries it is easy to find that in the CM frame we have:

$$n_{++}(q_0) = n_{--}(q_0) == \frac{1}{e^{\frac{\beta_1+\beta_2}{2}|q_0|} - 1} \equiv \tilde{n}(|q_0|\frac{\beta_1+\beta_2}{2}). \tag{A.23}$$

Computing n_{ij} for $i \neq j$ we must take into account that we have one additional particle:

$$n_{+-}(q_0) == \Theta(q_0)(1 + \tilde{n}(q_0\beta_1)) + \Theta(-q_0)\tilde{n}(-q_0\beta_1) \tag{A.24}$$

and

$$n_{-+}(q_0) = \Theta(q_0)\tilde{n}(q_0\beta_2) + \Theta(-q_0)(1 + \tilde{n}(-q_0\beta_2)). \tag{A.25}$$

Using (A.23), (A.24) and (A.25), and the definition (A.18) we find the Green functions:

$$G_{i,j}(x - x', (\beta)) = \int \frac{d^4q}{(2\pi)^4} e^{iq(x-x')} \tilde{G}_{ij}(q, (\beta))$$

where

$$i\tilde{G}_{ij}(q, (\beta)) = \begin{pmatrix} \frac{i}{q^2-m^2+i\epsilon} & 0 \\ 0 & -\frac{i}{q^2-m^2-i\epsilon} \end{pmatrix} +$$

$$+2\pi\delta(q^2 - m^2) \begin{pmatrix} \tilde{n}(\frac{\beta_1+\beta_2}{2}|q_0|) & \tilde{n}(\beta_2|q_0|)a_+(\beta_2) \\ \tilde{n}(\beta_1|q_0|)a_-(\beta_1) & \tilde{n}(\frac{\beta_1+\beta_2}{2}|q_0|) \end{pmatrix} \tag{A.26}$$

and

$$a_\pm(\beta) = -e^{\frac{\beta}{2}(|q_0|\pm q_0)}.$$

The corresponding generating functional has the standard form:

$$\rho_{cp}(j_\pm) = \exp\{-iV(-i\hat{j}_+) + iV(-i\hat{j}_-)\} \times$$
$$\times \exp\{\frac{i}{2}\int dxdx' j_i(x)G_{ij}(x - x', (\beta))j_j(x')\} \tag{A.27}$$

where the summation over repeated indexes is assumed.

Inserting (A.27) in the equation of state (A.14) we can find that $\beta_1 = \beta_2 = \beta(E)$. If $\beta(E)$ is a 'good' parameter then $G_{ij}(x - x'; \beta)$ coincides with the Green functions of the real-time finite-temperature field theory and the KMS boundary condition:

$$G_{+-}(t - t') = G_{-+}(t - t' - i\beta), \quad G_{-+}(t - t') = G_{+-}(t - t' + i\beta), \tag{A.28}$$

is restored. Equation (A.28) can be deduced from (A.26) by direct calculations.

APPENDIX B: LOCAL-TEMPERATURE S-MATRIX FORMALISM

B1 *VACUUM BOUNDARY CONDITION*

To describe this situation we decompose δ-functions in (A.4) on the product of $(N+1)$ δ-functions:

$$\delta^{(4)}(P-\sum_{k=1}^{n}q_k)=\int\prod_{\nu=1}^{N}\{dQ_\nu\delta(Q_\nu-\sum_{k=1}^{n_\nu}q_{k,\nu})\}\delta^{(4)}(P-\sum_{\nu=1}^{N}Q_\nu),$$

where $q_{k,\nu}$ is the momentum of the k-th in-going particle in the ν-th cell and Q_ν is the total 4-momenta of n_ν in-going particles in this cell, $\nu=1,2,...,N$. The same decomposition will be used for the second δ-function in (A.4). We must take into account the multinomial character of particle decomposition on N groups. This will give the coefficient:

$$\frac{n!}{n_1!\cdots n_N!}\delta_K(n-\sum_{\nu=1}^{N}n_\nu)\frac{m!}{m_1!\cdots m_N!}\delta_K(m-\sum_{\nu=1}^{N}m_\nu),$$

where δ_K is the Kronecker's δ-function.

As a result, the quantity

$$r((Q)_N,(P)_N)=\sum_{(n.m)}\int|a_{(n,m)}|^2\times$$
$$\times\prod_{\nu=1}^{N}\{\prod_{k=1}^{n_\nu}\frac{d\omega(q_{k,\nu})}{n_\nu!}\delta^{(4)}(Q_\nu-\sum_{k=1}^{n_\nu}q_{k,\nu})\prod_{k=1}^{m_\nu}\frac{d\omega(p_{k,\nu})}{m_\nu!}\delta^{(4)}(P_\nu-\sum_{k=1}^{m_\nu}p_{k,\nu})\}\qquad(B.1)$$

describes a probability to measure in the ν-th cell the fluxes of in-going particles with total 4-momentum Q_ν and of out-going particles with the total 4-momentum P_ν. The sequence of these two measurements is not fixed.

The Fourier transformation of δ-functions in (B.1) gives:

$$r((Q)_N,(P)_N)=\int\prod_{k=1}^{N}\frac{d^4\alpha_{1,\nu}}{(2\pi)^4}\frac{d^4\alpha_{2,\nu}}{(2\pi)^4}e^{i\sum_{\nu=1}^{N}(Q_\nu\alpha_{1,\nu}+P_\nu\alpha_{2,\nu})}\rho((\alpha_1)_N,(\alpha_2)_N),$$

where

$$\rho((\alpha_1)_N,(\alpha_2)_N)=\rho(\alpha_{1,1},\alpha_{1,2}...,\alpha_{1,N};\alpha_{2,1},\alpha_{2,2},...,\alpha_{2,N})$$

has the form:

$$\rho((\alpha_1)_N,(\alpha_2)_N)=\int\prod_{\nu=1}^{N}\{\prod_{k=1}^{n_\nu}\frac{d\omega(q_{k,\nu})}{n_\nu!}e^{-i\alpha_{1,\nu}q_{k,\nu}}\times$$
$$\times\prod_{k=1}^{m_\nu}\frac{d\omega(p_{k,\nu})}{m_\nu!}e^{-i\alpha_{2,\nu}p_{k.\nu}}\}|a_{(n,m)}|^2.\qquad(B.2)$$

Inserting (A.1) into (B.2) we find:

$$\rho((\alpha_-)_N,(\alpha_+)_N)=\exp\{i\sum_{\nu=1}^{N}\int dxdx'[\hat\phi_+(x)D_{+-}(x-x';\alpha_{2,\nu})\hat\phi_-(x')-$$
$$-\hat\phi_-(x)D_{-+}(x-x';\alpha_{1,\nu})\hat\phi_+(x')]\}Z(\phi_+)Z^*(\phi_-),\qquad(B.3)$$

where ϕ_- is defined on the complex conjugate contour $C_-:t\to t-i\varepsilon$ and $D_{+-}(x-x';\alpha)$, $D_{-+}(x-x';\alpha)$ are the positive and negative frequency correlation functions, respectively.

We must integrate over sets $(Q)_N$ and $(P)_N$ if the distribution of flux momenta over cells is not fixed. As a result,

$$r(P) = \int D^4\alpha_1(P) d^4\alpha_2(P) \rho((\alpha_1)_N, (\alpha_2)_N), \tag{B.4}$$

where the differential measure

$$D^4\alpha(P) = \prod_{\nu=1}^{N} \frac{d^4\alpha_\nu}{(2\pi)^4} K(P, (\alpha)_N)$$

takes into account the energy-momentum conservation laws:

$$K(P, (\alpha)_N) = \int \prod_{\nu=1}^{N} d^4Q_\nu e^{i\sum_{\nu=1}^{N}\alpha_\nu Q_\nu} \delta^{(4)}(P - \sum_{\nu=1}^{N} Q_\nu).$$

The explicit integration gives

$$K(P, (\alpha)_N) \sim \prod_{\nu=1}^{N} \delta^{(3)}(\alpha - \alpha_\nu),$$

where $\vec{\alpha}$ is the centre of mass (CM) 3-vector.

To simplify the consideration let us choose the CM frame and put $\alpha = (-i\beta, \vec{0})$. As a result,

$$K(E, (\beta)_N) = \int_0^\infty \prod_{\nu=1}^{N} dE_\nu e^{\sum_{\nu=1}^{N}\beta_\nu E_\nu} \delta(E - \sum_{\nu=1}^{N} E_\nu).$$

Correspondingly, in the CM frame,

$$r(E) = \int D\beta_1(E) D\beta_2(E) \rho((\beta_1)_N, (\beta_2)_N),$$

where

$$D\beta(E) = \prod_{\nu=1}^{N} \frac{d\beta_\nu}{2\pi i} K(E, (\beta)_N)$$

and $\rho((\beta)_N)$ was defined in (B.3) with $\alpha_{k,\nu} = (-i\beta_{k,\nu}, \vec{0})$, $Re\beta_{k,\nu} > 0$, $k = 1, 2$.

We will calculate integrals over β_k using the stationary phase method. The equations for the most probable values of β_k:

$$-\frac{1}{K(E, (\beta_k)_N)} \frac{\partial}{\partial\beta_{k,\nu}} K(E, (\beta_k)_N) = \frac{1}{R((\beta_1)_N)} \frac{\partial}{\partial\beta_{k,\nu}} R((\beta)_N), \quad k = 1, 2, \tag{B.5}$$

always have the unique positive solutions $\tilde{\beta}_{k,\nu}(E)$. We propose that the fluctuations of β_k near $\tilde{\beta}_k$ are small, i.e. are Gaussian. This is the basis of the local-equilibrium hypothesis [26]. In this case $1/\tilde{\beta}_{1,\nu}$ is the temperature in the initial state in the measurement cell ν and $1/\tilde{\beta}_{2,\nu}$ is the temperature of the final state in the ν-th measurement cell.

The last formulation (B.4) implies that the 4-momenta $(Q)_N$ and $(P)_N$ can not be measured. It is possible to consider another formulation also. For instance, we can suppose that the initial set $(Q)_N$ is fixed (measured) but $(P)_N$ is not. In this case we will have a mixed experiment: $\tilde{\beta}_{1,\nu}$ is defined by the equation:

$$E_\nu = -\frac{1}{R} \frac{\partial}{\partial\beta_{1,\nu}} R$$

and $\tilde{\beta}_{2,\nu}$ is defined by the second equation in (B.5).

Considering the limit $N \to \infty$ the dimension of cells tends to zero. In this case we are forced by the quantum uncertainty principle to propose that the 4-momenta sets (Q) and (P) are not fixed. This formulation becomes pure thermodynamical: we must assume that (β_i) and (β_f) are measurable quantities. For instance, we can fix (β_i) and try to find (β_f) as the function of the total energy E and the functional of (β_i). In this case Eqs. (B.5) become the functional equations.

In the considered microcanonical description the finiteness of the temperature does not touch the quantization mechanism. Really, one can see from (B.3) that all thermodynamical information is confined in the operator exponent

$$e^{\hat{N}(\phi_i^*\phi_j)} = \prod_{\nu}\prod_{i\neq j} e^{i\int \hat{\phi}_i D_{ij}\hat{\phi}_j}$$

the expansion of which describes the environment, and the 'mechanical' perturbations are described by the amplitude $Z(\phi)$. This factorization was achieved by the introduction of an auxiliary field ϕ and it is independent from the choice of boundary conditions, i.e. from the choice of considered systems environment.

B2 *CLOSED PATH BOUNDARY CONDITIONS*

The developed formalism allows one to introduce the more general 'closed-path' boundary conditions. The presence of the external black-body radiation flow will reorganize only $\exp\{\hat{N}(\phi_i^*\phi_j)\}$, the differential operator, and the new generating functional ρ_{cp} has the form:

$$\rho_{cp}(\alpha_1,\alpha_2) = e^{\hat{N}(\phi_i^*\phi_j)}\rho_0(\phi_\pm).$$

Introducing the cells we will find that

$$\hat{N}(\phi_i^*\phi_j) = \int drdy\hat{\phi}_i(r+y/2)\tilde{n}_{ij}(Y,y)\hat{\phi}_j(r-y/2),$$

where the occupation number $\tilde{n}_{ij}$ carries the cells index r:

$$\tilde{n}_{ij}(r,y) = \int d\omega(q)e^{iqy}n_{ij}(r,q)$$

and $(q_0 = \epsilon(q))$

$$n_{++}(r,q_0) = n_{--}(r,q_0) = \tilde{n}(r,(\beta_1+\beta_2)|q_0|/2) = \frac{1}{e^{(\beta_1+\beta_2)(r)|q_0|/2}-1},$$

$$n_{+-}(r,q_0) = \Theta(q_0)(1+\tilde{n}(r,\beta_2 q_0)) + \Theta(-q_0)\tilde{n}(r,-\beta_1 q_0),$$

$$n_{-+}(r,q_0) = n_{+-}(r,-q_0).$$

For simplicity the CM system was used.

Calculating ρ_0 perturbatively we will find that

$$\rho_{cp}(\beta) = \exp\{-iV(-i\hat{j}_+) + iV(-i\hat{j}_-)\}\times$$
$$\exp\{i\int drdy[\hat{j}_i(r+y/2)G_{ij}(y,(\beta(r))\hat{j}_j(r-y/2)\} \tag{B.6}$$

where, using the matrix notations,

$$iG(q,(\beta(r))) = \begin{pmatrix} \frac{i}{q^2-m^2+i\varepsilon} & 0 \\ 0 & -\frac{i}{q^2-m^2-i\varepsilon}\end{pmatrix} +$$

$$+2\pi\delta(q^2-m^2)\begin{pmatrix} n(\frac{(\beta_1+\beta_2)(r)}{2}|q_0|) & n(\beta_1(r)|q_0|)a_+(\beta_1) \\ n(\beta_2(r)|q_0|)a_-(\beta_2) & n(\frac{(\beta_1+\beta_2)(r)}{2}|q_0|) \end{pmatrix}, \tag{B.7}$$

and

$$a_\pm(\beta) = -e^{\beta(|q_0|\pm q_0)/2}. \tag{B.8}$$

Formally these Green functions obey the standard equations in the y space:

$$(\partial^2 - m^2)_y G_{ii} = \delta(y),$$

$$(\partial^2 - m^2)_y G_{ij} = 0, \;\; i \neq j$$

since $\Phi(\sigma_\infty) \neq 0$ reflects the mass-shell particles. But the boundary conditions for these equations are not evident.

It should be underlined that in our consideration r is the coordinate of $measurement$, i.e. r is like the calorimeter cells coordinate and there is no necessity to divide the interaction region of QGP into domains (cells). This means that L must be smaller then the typical range of fluctuations of QGP. But, on other hand, L can not be arbitrarily small since this will lead to the assumption of the $local$ factorization property of correlators, i.e. to the absence of interactions.

So, changing $\beta \to \beta(r)$ we should assume that $\beta_{i(f)}(r)$ and $z_{+-(-+)}(r,k)$ are constants on the interval L. This prescription adopts the Wigner functions formalism for the case of high multiplicities. It describes the temperature fluctuations larger than L and averages the fluctuations smaller than L leading to the absence, on average, of 'non- Gaussian' fluctuations.

It is the typical 'calorimetric' measurement, since in a dominant number of calorimeter cells the measured mean values of energy, with exponential accuracy, are the 'good' parameters $\sim 1/\beta_2(r,E)$. We will assume that the dimension of calorimeter cells $L << L_{cr}$, where L_{cr} is the dimension of characteristic fluctuations at a given n. In deep asymptotic over n we must have $L_{cr} \to \infty$. This consideration shows that the offered experiment with calorimeter as the measuring device of particle energies is sufficiently informative in the high-multiplicities domain.

APPENDIX C: WIGNER FUNCTIONS S-MATRIX FORMALISM

C1 *WIGNER FUNCTIONS FORMALISM*

We shall use the Wigner functions formalism in the Carrusers–Zachariasen formulation [27]. For the sake of generality the m into n particles transition will be considered. This will allow us to take the heavy ion–ion collisions into consideration.

In the previous section the generating functional $R((\beta)_N)$ was calculated by means of dividing the 'measuring device' (calorimeter) on the N cells. It was assumed that the dimension of device cells tends to zero ($N \to \infty$). Now we shall specify the cell coordinates using Wigner's description.

Let us introduce the distribution function F_n which defines the probability to find n particles with definite momentum and with arbitrary coordinates. These probabilities (cross sections) are usually measured in particle physics. The corresponding Fourier-transformed generating functional can be deduced from (B.3):

$$F(z,(\beta_+)_N,(\beta_-)_N) = \prod_{\nu=1}^{N}\prod_{i\neq j} e^{\int d\omega(q)\hat{\phi}_i^*(q)e^{-\beta_{j,\nu}\epsilon(q)}\hat{\phi}_j(q)z_{ij}^\nu(q)} \times$$

$$\times Z(\phi_+)Z^*(\phi_-). \tag{C.1}$$

The variation of F over $z_{ij}^\nu(q)$ generates corresponding distribution functions. One can interpret $z_{ij}^\nu(q)$ as the local activity: the logarithm of $z_{ij}^\nu(q)$ is conjugate to the particles number in the cell ν with

momentum q for the initial ($ij = 21$) or final ($ij = 12$) states. Note that $z^{\nu}_{ij}(q)\hat{\phi}^{*}_{i}(q)\hat{\phi}_{j}(q)$ can be considered as the operator of activity.

The Boltzmann factor $e^{-\beta_{i,\nu}\epsilon(q)}$ can be interpreted as the probability to find a particle with the energy $\epsilon(q)$ in the final state ($i = 2$) and in the initial state ($i = 1$). The total probability, i.e. the process of creation and further absorption of n particles, is defined by the multiplication of these factors.

The generating functional (C.1) is normalized as follows:

$$F(z = 1, (\beta)) = R((\beta)), \tag{C.2}$$

$$F(z = 0, (\beta)) = |Z(0)|^2 = \rho_0(\phi_{\pm})|_{\phi_{\pm}=0}$$

Where

$$\rho_0(\phi_{\pm}) = Z(\phi_{+})Z^{*}(\phi_{-})$$

is the 'probability' of the vacuum into vacuum transition in the presence of auxiliary fields $\phi_{\pm}$. The one-particle distribution function

$$F_1((\beta_1)_N, (\beta_2)_N; q) = \frac{\delta}{\delta z^{\nu}_{ij}(q)} F|_{z=0} =$$
$$= \{\hat{\phi}^{*}_{i}(q)e^{-\beta^{\nu}_{i}\epsilon(q)/2}\}\{\hat{\phi}_{j}(q)e^{-\beta^{\nu}_{i}\epsilon(q)/2}\}\rho_0(\phi_{\pm}) \tag{C.3}$$

describes the probability to find one particle in the vacuum.

So,

$$F_1((\beta_1)_N, (\beta_2)_N; q) = \int dxdx' e^{iq(x-x')}e^{-\beta_{i,\nu}\epsilon(q)}\}\hat{\phi}_{i}(x)\hat{\phi}_{j}(x')\rho_0(\phi_{\pm}) =$$
$$= \int dr\{dy e^{iqy}e^{-\beta_{i,\nu}\epsilon(q)}\}\hat{\phi}_{i}(r + y/2)\hat{\phi}_{j}(r - y/2)\rho_0(\phi_{\pm})\}. \tag{C.4}$$

Using this definition we introduce the one-particle Wigner function W_1 [27]:

$$F_1((\beta_1)_N, (\beta_2)_N; q) == \int dr W_1((\beta_1)_N, (\beta_2)_N; r, q).$$

So,

$$W_1((\beta_1)_N, (\beta_2)_N; r, q) = \int dy e^{iqy}e^{-\beta_{i,\nu}\epsilon(q)}\hat{\phi}_{i}(r + y/2)\hat{\phi}_{j}(r - y/2)\rho_0(\phi_{\pm}).$$

This distribution function describes the probability to find in the vacuum at point r in cell ν a particle with momentum q.

Since the choice of the device coordinates is in our hands it is natural to adjust the cell coordinate to the coordinate of the measurement r:

$$W_1((\beta_1)_N, (\beta_2)_N; r, q) = \int dy e^{iqy}e^{-\beta_{i}(r)\epsilon(q)}\}\hat{\phi}_{i}(r + y/2)\hat{\phi}_{j}(r - y/2)\rho_0(\phi_{\pm}).$$

This choice of the device coordinates leads to the following generating functional:

$$F(z, \beta) = \exp\{i \int dydr[\hat{\phi}_{+}(r + y/2)D_{+-}(y; \beta_2(r), z)\hat{\phi}_{-}(r - y/2) -$$
$$-\hat{\phi}_{-}(r + y/2)D_{-+}(y; \beta_1(r), z)\hat{\phi}_{+}(r - y/2)]\}\rho_0(\phi_{\pm}), \tag{C.5}$$

where

$$D_{+-}(y; \beta(r), z) = -i \int d\omega(q) z_{+-}(r, q)e^{iqy}e^{-\beta(r)\epsilon(q)},$$

$$D_{-+}(y;\beta(r),z) = i\int d\omega(q) z_{-+}(r,q) e^{-iqy} e^{-\beta(r)\epsilon(q)}$$

are the modified positive and negative correlation functions.

The inclusive, partial, distribution functions are familiar in particle physics. These functions describe the distributions in the presence of the arbitrary number of other particles. For instance, the one-particle partial distribution function

$$P_{ij}(r,q;(\beta)) = \frac{\delta}{\delta z_{ij}(r,q)} F(z,(\beta))|_{z=1} =$$
$$= \frac{e^{-\beta_i(r)\epsilon(q)}}{(2\pi)^3\epsilon(q)} \int dy e^{iqy} \hat{\phi}_i(r+y/2)\hat{\phi}_j(r-y/2)\rho(\phi_\pm,(\beta)), \tag{C.6}$$

where Eq. (C.2) was used.

The mean multiplicity $n_{ij}(r,q)$ of particles in the infinitesimal cell Y with momentum q is

$$n_{ij}(r,q) = \int dq \frac{\delta}{\delta z_{ij}(r,q)} \ln F(z,(\beta))|_{z=1}.$$

If the interactions among fields are switched out we can find that (omitting indexes):

$$n(Y,q_0) = \frac{1}{e^{\beta(r)q_0} - 1}, \quad q_0 = \epsilon(q) > 0.$$

This is the mean multiplicity of the black-body radiation.

C2 *LIOUVILLE EQUATION FOR THE WIGNER FUNCTION*

Let us consider:

$$W_1((\beta_f)_N,(\beta_i)_N;Y,q) = \int dy e^{iqy} e^{-\beta_k(Y)\varepsilon(q)} \hat{\phi}_k(Y+y/2)\hat{\phi}_j(Y-y/2) R_0(\phi_\pm).$$

We would like to investigate under what conditions W_1 obey the Liouville equation.

The functional integral representation for W_1 has a form:

$$W_1(\beta;Y,q) = \int dy e^{iqy} e^{-\beta_k(Y)\varepsilon(q)} \hat{\phi}_k(Y+y/2)\hat{\phi}_j(Y-y/2)\times$$
$$\times \int D\Phi_+ D\Phi_- e^{iS^0_{C_+(t_{in})}(\Phi_+) - iV_{C_+(t_{in})}(\Phi_+ + \phi_f) - iS^0_{C_-(t_{in})}(\Phi_-) + iV_{C_-(t_{in})}(\Phi_- - \phi_i)} =$$
$$= \int dy e^{iqy} e^{-\beta_k(Y)\varepsilon(q)} \int D\Phi_+ D\Phi_- V'_{C_+(t_{in})}(\Phi_+; Y+y/2) V'_{C_-(t_{in})}(\Phi_-; Y-y/2)\times$$
$$\times e^{iS_{C_+(t_{in})}(\Phi_+) - iS_{C_-(t_{in})}(\Phi_-)}, \tag{C.7}$$

where $S^0_{C_\pm(t_{in})}$ is the free part of the total action,

$$S_{C_\pm(t_{in})}(\Phi_\pm) = S^0_{C_\pm(t_{in})}(\Phi_\pm) - V_{C_\pm(t_{in})}(\Phi_\pm),$$

and the Mills time contour

$$C_\pm(t_{in}):\ t \to \pm i\epsilon,\ \epsilon \to +0,\ t_{in} \le t \le +\infty,\ t_{in} \to -\infty \tag{C.8}$$

was introduced. We should use the closed-path boundary condition:

$$\Phi_\pm(\sigma_{t_{in}}) = \Phi(\sigma_{t_{in}}), \tag{C.9}$$

where the hypersurface $\sigma_{t_{in}}$ crosses the point t_{in}. It should be stressed that the integration over the 'turning-point' field $\Phi(\sigma_{t_{in}})$ must be performed. Notice that the closed-path boundary condition only allows at finite t_{in} to be saved from the unlike 'surface terms'.

The representation (C.7) contains the vertices

$$V'_{C_-(t_{in})}(\Phi_\pm; Y \pm y/2) = \frac{\delta}{\delta\Phi_\pm(Y \pm y/2)} V_{C_-(t_{in})}(\Phi_\pm).$$

Notice that there is no necessity to cut the integral over y at $y = t_{in}$ since the action of the operators $)\hat{\phi}_j(Y \pm y/2)$ ends at the time $(Y \pm y/2)_0 = t_{in}$. Therefore, $W_1(\beta; Y, q)$ exist for the time interval $Y_0 < t_{in}$. The t_{in} dependence of $D\Phi_\pm$ on t_{in} is not important since we always can add the (infinite) integration over

$$\prod_{t=-\infty}^{t_{in}} d\Phi(t)$$

assuming that this infinity may be cancelled by the normalization factor.

It is known that the double functional integral (C.7) is defined on the δ-like Dirac measure [7]. The result looks as follows:

$$W_1(\beta; Y, q) = \int dy e^{iqy} e^{-\beta_k(Y)\varepsilon(q)} e^{-i\mathbf{K}(je)} \times$$

$$\times \int DM(\Phi) V'_{C_+(t_{in})}(\Phi + e; Y + y/2) V'_{C_-(t_{in})}(\Phi - e; Y - y/2) e^{-iU_C(\Phi; e)}. \tag{C.10}$$

Expanding $\exp\{-i\mathbf{K}(je)\}$ over the operator

$$2\mathbf{K}(je) = \int dx \frac{\delta}{\delta j(x)} \frac{\delta}{\delta e(x)}, \tag{C.11}$$

we will obtain the ordinary perturbation theory. Notice that the operator $\mathbf{K}(je)$ is t_{in} independent.

The functional integral (C.10) is defined on the measure:

$$DM(\Phi) = \prod_x{}' d\Phi(x) \delta\left(\frac{\delta S(\Phi)}{\delta\Phi(x)} - j(x)\right), \tag{C.12}$$

where the prime means that the functional δ-function does not include the time end-point $x_0 = t_{in}$.

At the end of definitions, the functional $U_{C_+}(\Phi; \phi, e)$ describes interactions [7]:

$$U_{C_+}(\Phi; e) = ... \tag{C.13}$$

The explicit form of it is not important for us.

Deriving the Liouville equation, the dynamics should be described in the phase space. It is easy to see that the measure (C.12) has the following form in the phase space:

$$DM(\Phi) = \prod_x d\Phi(x) d\mathrm{P}(x) \delta\left(\dot{\Phi} - \frac{\delta H_j(\Phi, \mathrm{P})}{\delta \mathrm{P}(x)}\right) \delta\left(\dot{\mathrm{P}} + \frac{\delta H_j(\Phi, \mathrm{P})}{\delta\Phi(x)}\right), \tag{C.14}$$

where the Hamiltonian

$$H_j(\Phi, \mathrm{P}) = \int d^3x \left\{\frac{1}{2}\mathrm{P}^2 + \frac{1}{2}(\nabla\Phi)^2 + v(\Phi) - j\Phi\right\} \tag{C.15}$$

explicitly depends on the produced quantum perturbation force $j(x)$ and $v(\Phi)$ is the potential term. The transition from (C.12) to (C.14) may raise a doubt caused by a possible symmetry of the problem under

consideration. To avoid this ambiguity, one may consider this transition as the introduction of the first order formalism.

Notice now that the equalities:

$$\dot{\Phi} = \frac{\delta H_j(\Phi, \mathrm{P})}{\delta \mathrm{P}(x)}, \;\; \dot{\mathrm{P}} = -\frac{\delta H_j(\Phi, \mathrm{P})}{\delta \Phi(x)} \tag{C.16}$$

can not fix the boundary values Φ_0 and P_0. For this reason one may omit the prime in the measure (C.14), i.e. including the boundary value of $t = t_{in}$ in the Dirac measure.

Now it is important to note that, following our definition,

$$\int \prod_x d\Phi(x)\delta(\dot{\Phi}(x)) = \int d\Phi(t_{in}) = \int d\Phi_0.$$

Therefore, the boundary values (Φ_0, P_0) stay undefined by our functional δ-functions and, generally speaking, the integration over them is assumed:

$$W_1(\beta; Y, q, t_i n) = \int d\Phi_0 d\mathrm{P}_0 W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0). \tag{C.17}$$

The Liouville equation exists just for $W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0)$.

Indeed, let us calculate the total derivative over t_{in}:

$$\frac{d}{dt_{in}} W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0) = \frac{\partial W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0)}{\partial \Phi_0} \dot{\Phi}_0 + \frac{\partial W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0)}{\partial \mathrm{P}_0} \dot{\mathrm{P}}_0. \tag{C.18}$$

But having the measure (C.14), one may write that

$$\frac{d}{dt_{in}} W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0) = e^{-i\mathbf{K}(je)} e^{-iU_C(\Phi;0,e)} \left\{ W_1(\beta; Y, q, \Phi_0, \mathrm{P}_0), H_j(\Phi_0, \mathrm{P}_0) \right\}, \tag{C.19}$$

where the Poisson bracket

$$\{W_1, H_j\} = \frac{\partial W_1}{\partial \Phi} \frac{\partial H_j}{\partial \mathrm{P}} - \frac{\partial W_1}{\partial \mathrm{P}} \frac{\partial H_j}{\partial \Phi}.$$

Notice the quantum character of this equation: the r.h.s. contains the operator of quantum perturbations $\mathbf{K}(je)$. It acts on j in the Hamiltonian $H_j = H_j(\Phi_0, \mathrm{P}_0)$ and e in the interaction functional $U_C(\Phi; 0, e)$. Notice, all quantities are defined at the time moment t_{in}.

APPENDIX D: NONSTATIONARY STATISTICAL OPERATOR

One can not expect the evident connection between the above considered S-matrix (microcanonical) and Zubarev's [26] approaches. The reason is the introduction into Zubarev's formalism interaction of a heat bath, external to the system under consideration. This interaction is crucial for the definition of nonstationary statistical operator (NSL) to explain the trend to a maximal-entropy state, starting the evolution from the local-equilibrium state[2].

Therefore, in Zubarev's theory the local-equilibrium state was chosen as the boundary condition. It is assumed that in the suitably defined cells of the $system$ at a given temperature distribution $T(\vec{x}, t) = 1/\beta(\vec{x}, t)$, where $(\vec{x}, t)$ is the index of the cell, the entropy is maximal. The corresponding nonequilibrium statistical operator

$$\rho_z \sim e^{-\int d^3x \beta(\vec{x},t) T_{00}} \tag{D.1}$$

[2]This condition is not necessary in the S-matrix formalism since it is 'dynamical' by its nature, i.e. includes the notion of initial and final states as the boundary conditions.

describes the evolution of a system in the time scale t. Here $T_{\mu\nu}$ is the energy-momentum tensor. It is assumed that the system 'follows' $\beta(\vec{x},t)$ evolution and the local temperature $T(\vec{x},t)$ is defined as the external parameter which is the regulator of systems dynamics. For this purpose the special $i\varepsilon$-prescription was introduced (it was not shown in (D.1) [26]. It brings the interaction with the heat bath.

The KMS periodic boundary condition can not be applied for the nonstationary temperature distribution and by this reason the decomposition:

$$\beta(\vec{x},t) = \beta_0 + \beta_1(\vec{x},t) \tag{D.2}$$

was offered [28]. Here β_0 is the constant and the inequality

$$\beta_0 >> |\beta_1(\vec{x},t)|$$

is assumed. Then,

$$\rho_z \sim e^{-\beta_0(H_0+V+B)} \tag{D.3}$$

where H_0 is the free part of the Hamiltonian, V describes the interactions and the linear over β_1/β_0 term B is connected with the deviation of the temperature from the 'equilibrium' value $1/\beta_0$. The presence of B-perturbations creates the 'thermal' flows in the system to explain the increasing entropy. Considering V and B as the perturbations, one can calculate the observables averaging over equilibrium states, i.e. adopting the KMS boundary condition. Using the standard terminology one can consider V as the 'mechanical' and B as the 'thermal' perturbations.

The quantization problem of the operator (D.3) is connected to the definition of the space-time sequence of mechanical (V) and thermal (B) excitations. It is necessary since the mechanical excitations have the influence on the thermal ones and vice versa. It was assumed in Ref. [28] that V and B are commuting operators, i.e. the sequence of V- and B-perturbations is not sufficient. The corresponding generating functional has the form [28]:

$$Z(j) = \exp\{-i\int_{C_\beta} d^4x(V(-i\hat{j}(x)) + \frac{\beta_i(\mathbf{x},\tau)}{\beta_0}T_{00}[-i\hat{j}(x)] -$$

$$\int_{-\infty}^{0} dt_1 \frac{\beta_i(\mathbf{x},\tau+t_1)}{\beta_0}T_{00}[-i\hat{j}(\mathbf{x},x_0t_1)])\}Tr(e^{-\beta_0 H_0}T_C e^{i\int_C d^4y j(y)\Phi(y)}),$$

where τ is the measurement time.

It is evident that this solution leads to renormalization by the interactions with the external field $\beta(\vec{x},t)$ even without interactions among fundamental fields Φ. The source of these renormalizations is the kinetic term in the energy-momentum tensor T_{00}, i.e. follows from 'thermal' interactions with the external heat bath. Note the absence of renormalizations in the S-matrix formalism, see, for instance (A.27), where the interactions are generated by V perturbations only.

This formulation with $\beta(\vec{x},t)$ as the external field reminds one of the old, firstly quantized, field theory in which matter is quantized but fields are not. It is well known that consistent quantum field theory requires the second quantization. Following this analogy, if we want to take into account consistently the reciprocal influence of V- and B-perturbations, the field $\beta(\vec{x},t)$ must be fundamental, i.e. must be quantized (and in this case the assumption of Ref. [28] becomes true). But it is evidently the wrong idea in the canonical Gibbs formalism. So, as in the firstly quantized theory, the theory with the operator (D.1) must have the restricted range of validity [26].

REFERENCES

[1] E. Fermi, *Progr. Theor. Phys.* **4** (1950) 570, *Phys. Rev.* **81** (1950) 115, *ibid.* **92** (1953) 452;
L.D. Landau, *Izv. AN SSSR,* **17** (1953) 85.

[2] F. Becattini, *et al.*, hep-ph/0002267; hep-ph/00110221; hep-ph/0206203;
P. Braun-Munzinger, *et al.*, nucl-th/9903010;
U. Heinz and P.F. Kolb, hep-ph/0204061;
U. Henz, *Nucl. Phys.* **A661** (1999) 140c;
P. Braun-Munzinger, *et al.*, hep-ph/0105229;
H. Oeschler, nucl-ex/0011007;
Zhong-Dao Lu, hep-ph/0207029;
R. Baier *et al.*, hep-ph/0204211;
J.B. Elliot *et al.*, *Phys. Rev. Lett.* **85** (2000) 1194;
C. Tsallis, *Lect. Notes in Phys.* **LNP 560** (2000);
G.A. Kozlov, *New J. Phys.* **4** (2002) 23;
D. Kharzeev, hep-ph/0204015; E. Shuryak, hep-ph/0205031;
I.M. Dremin and V.A. Nechitailo, hep-ph/0207068;
L. Gutay *et al.*, E-735 Coll. (FNAL), ISMD-02.

[3] Rev. Part. Phys., *Phys. Rev.* **D66** (2002) 258.

[4] M. Kac, *Probability and Related Topics* (Interscience, London, 1957).

[5] E. Fermi, J. Pasta, and S. Ulam, *Studies in Nonlinear Problems,* I. Los Alamos report LA 1940 (1955). Reproduced in Nonlinear Wave Motion, Ed. A. C. Newell. (Amer. Math. Soc., Providence, RI, 1974).

[6] M.D. Kruskal and N.J. Zabuski, *Princeton Plasma Physics Lab. Annual Rep.*, MATT-Q-21 (Princeton, NJ, 1963);
M.D. Kruskal, *Proc. IBM Scient. Comp. Symp.* (White Planes, NY, 1965); V.E. Zakharov, *JETP* **65** (1973) 219.

[7] J. Manjavidze and A. Sissakian, *Phys. Rep.* **346** (2000) 1;
J. Manjavidze, *Elem. Part. and Atom. Nucl.* **16** (1985) 101.

[8] R.P. Feynman, *Phys. Rev. Lett.* **23** (1969) 1415.

[9] R. Kubo, *J. Phys. Soc. Japan* **12** (1957) 570.

[10] L.D. Landau, E.M. Lifshitz and L.P. Pitaevskii, *Statistical Physics* (Pergamon Press, Oxford, 1980).

[11] N.N. Bogolyubov, *Studies in Statistical Mechanics*, eds. J. DeBoer and G.E. Uhlenbeck (North-Holland, Amsterdam, 1962).

[12] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1965).

[13] L. Landau and R. Peierls, *Z. Phys.* **69** (1931) 56.

[14] N.P. Landsman and Ch.G. van Weert, *Phys. Rep.* **145** (1987) 141.

[15] A.J. Niemi and G. Semenoff, *Ann. Phys. (NY)* **152** (1984) 105.

[16] R. Mills, *Propagators for Many-Particle Systems* (Gordon and Breach, New York, 1970).

[17] E. Byckling and K. Kajantie, *Particle Kinematics* (Wiley, London, 1973).

[18] J. Schwinger, *J. Math. Phys.* **A9** (1964) 2363; L. Keldysh, *Sov. Phys. JETP* **20** (1964) 1018.

[19] T. Matsubara, *Prog. Theor. Phys.* **14** (1955) 351.

[20] R. Haag, N. Hugengoltz and M. Winnink, *Commun. Math. Phys.* **5** (1967) 5.

[21] H. Chu and H. Umezawa, *Int. J. Mod. Phys.* **A9** (1994) 2363.

[22] M. Kac, G.E. Uhlenbeck and P.C. Hemmer, *J. Math. Phys.* **4** (1963) 216, *ibid.* **5** (1964) 60.

[23] J.D. Bjorken, *Acta Phys. Polon.* **B28** (1997) 2773, hep-ph/9712434.
[24] J. Schwinger, *Particles, Sources and Fields,* Vol.1 (Addison-Wesley, reading, MA, 1970).
[25] E. Calsetta and B.L. Hu, *Phys. Rev.* **D37** (1988) 2878.
[26] D.N. Zubarev, *Nonequilibrium Statistical Thermodynamics* (Consultants Bureau, NY, 1974).
[27] P. Carruthers and F. Zachariasen, *Phys. Rev.* **D13** (1986) 950.
[28] T. Bibilashvili and I. Pasiashvili, *Ann. Phys. (NY)* **220** (1992) 134.

ASTROPARTICLE PHYSICS

I. Tkachev
CERN, Geneva, Switzerland
and
Institute for Nuclear Research of the Russian Academy of Sciences, 117312, Moscow, Russia

Abstract
In this astroparticle lecture course I shall try to emphasize evidence of the new physics which we have in cosmological and astrophysical data. This includes support of the inflationary model, the necessity of dark energy and of non-baryonic dark matter, and the Grizen–Zatsepin–Kuzmin puzzle of ultra-high-energy cosmic rays.

1 INTRODUCTION

The purpose of these CERN physics school lectures is to review the evidence for new physics in cosmological and astrophysical data, and to give the minimal theoretical frameworks needed to understand and appreciate the evidence. Beyond any reasonable doubt, we have solid evidence for the new physics in these data. The strongest is the case for non-baryonic dark matter, followed by the case for dark energy. The possibility (though very speculative, since a consistent and working model has not been constructed yet) that the law of gravity should be changed instead is not excluded, but that would mean a new physics anyway. I will not engage in discussion of relevant particle physics model building, instead the reader is referred to lectures by G. Gabadadze at this school [1]. Other solid evidence for the new physics beyond the Standard Model is given by neutrino oscillations. I will not discuss this topic, it is covered in lectures by S. Petcov at this school [2]. The physics of cosmic rays is partially covered in lectures by A. Chilingarian [3], therefore, I restrict myself to the highest-energy part of the spectrum, which is related to the Grizen–Zatsepin–Kuzmin puzzle and, possibly, to a new physics.

There are many excellent reviews on the subject of cosmology and astroparticle physics, including lectures at previous CERN schools, for a recent one see Refs. [4–6]. I have tried to be complementary to these lectures as far as possible, so many additional details can be found there. Proceedings of these schools can be found at http://physicschool.web.cern.ch/ PhysicSchool. Owing to space and time limitations, I omit several very important traditional topics, most notably Big Bang nucleosynthesis (see, for example, the review in Ref. [7]) and baryogenesis (see, for example, the review in Ref. [8]). In the area covered, I have updated experimental results and resulting constraints, where applicable. The most important developments since the time of the previous school were the release of the first-year observations of Cosmic Microwave Background Radiation (CMBR) by the Wilkinson Microwave Anisotropy Probe (WMAP) [9], the first data release by the Sloan Digital Sky Survey (SDSS) of three-dimensional distribution of galaxies [10], and the release [11] of a statistically significant dataset of Supernovae Ia at large cosmological redshifts, $z > 1$, which provide the first conclusive evidence for cosmic deceleration that preceded the current epoch of cosmic acceleration. These are long-awaited cosmological data of unprecedented quality, and with their appearance cosmology has truly entered the golden era and became a precision science.

The plan of the lectures is as follows. In Section 2, I review the basics of cosmology: Friedman equations, Hubble expansion, cosmography. In Section 3, the Cosmic Microwave Background Radiation (CMBR) is discussed. In Section 4, I briefly review recent results on another cosmological probe—the large-scale distribution of galaxies. In Section 5, the evidence for the existence of dark energy is presented. Sections 6 and 6.3 review the evidence for dark matter, and particle physics models of non-baryonic matter are briefly considered. In Section 7, I review the basics of inflationary cosmology and

discuss support of the inflationary model by the CMBR data. In Section 8, the physics of ultra-high-energy cosmic rays is reviewed.

2 BASICS OF COSMOLOGY

2.1 Note on units and scales

Length. Astronomers measure distances in parsecs, which is about 3.1×10^{16} m or about 3.26 light years, and is comparable with the distance to the closest star. PARSEC is an abbreviation for the distance to a star with a semi-annual PARallax of 1 arc SECond. The distance from our Sun to the Galactic centre is 8 kiloparsecs, so the kpc is an appropriate unit when discussing galactic structure. The appropriate unit of extragalactic distance, however, is the megaparsec, or Mpc. The nearest large cluster of galaxies, the Virgo cluster, is at a distance of 20 Mpc. The size of the visible Universe is 4200 Mpc or 13.7 billion light years.

Energy. Usually astronomers measure energy in ergs. For example, the luminosity of our Sun is 4×10^{33} erg s^{-1}, while the luminosity of bright quasars reaches 10^{46} erg s^{-1}. A galaxy like our Milky Way contains 10^{11} stars, and there are 10^{11} galaxies in the visible part of the Universe. Particle physicists measure energy in electronvolts, 1 erg $= 6.2 \times 10^{11}$ eV, and usually choose units where the velocity of light and the Planck constant are set to unity, $c = 1$, $\hbar = 1$, which I also use, when convenient. In these units, for example, 1 Mpc $= 1.6 \times 10^{29}$ eV^{-1}.

2.2 Dynamical frameworks

Dynamics is provided by General Relativity—the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G \, T_{\mu\nu} \ , \tag{1}$$

where $T_{\mu\nu}$ is a stress energy tensor describing the distribution of mass in space, G is Newton's gravitational constant, and the curvature $R_{\mu\nu}$ is a complicated function of the metric and its first and second derivatives. Clearly, finding a general solution to a set of equations as complex as the Einstein field equations is a hopeless task. The problem is simplified greatly by considering mass distributions with special symmetries. The basic assumption underlying the construction of cosmological models is that of spatial homogeneity and isotropy. The most general space–time metric consistent with these symmetries is the Robertson–Walker metric:

$$ds^2 = dt^2 - a^2(t) \, \mathbf{dl}^2 \ , \tag{2}$$

where $a(t)$ is the dimensionless scale factor by which all distances vary as a function of cosmic time t. The scale factor contains all the dynamics of the Universe, while the vector product $\mathbf{dl}^2$ describes the geometry of space,

$$\mathbf{dl}^2 = \frac{dr^2}{1 - k\, r^2} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \ ,$$

which can be either Euclidian, or positively or negatively curved. For the spatial 3-dimensional curvature we find, explicitly

$${}^{(3)}R = \frac{6k}{a^2(t)} \quad \begin{cases} k = -1 & \text{Open} \\ k = 0 & \text{Flat} \\ k = +1 & \text{Closed} \ . \end{cases} \tag{3}$$

For example, the space with $k = +1$ can be thought of as a 3-dimensional sphere with a curvature being inversely proportional to the square of its radius. In this section we will model the matter content of the Universe as a perfect fluid with energy density ρ and pressure p, for which the stress-energy tensor in the

rest frame of the fluid is

$$T_\mu^{\ \nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix} \ . \tag{4}$$

With these assumptions the Einstein equations simplify to the Friedmann equations, which form the dynamical basis of cosmology

$$\frac{\dot{a}^2}{a^2} \ = \ \frac{8\pi G}{3}\,\rho \ - \ \frac{k}{a^2}\ , \tag{5}$$

$$\frac{\ddot{a}}{a} \ = \ -\,\frac{4\pi G}{3}\,(\rho + 3\,p)\ . \tag{6}$$

Let us have a look at the basic physics behind these equations.

1. Differentiating Eq. (5) and subtracting Eq. (6) we obtain

$$\frac{d\rho}{dt} \ + \ 3\,\frac{\dot{a}}{a}\,(\rho + p) = 0\ , \tag{7}$$

which is nothing but energy-momentum conservation,

$$T_\mu^{\ \nu}{}_{;\nu} = 0\ . \tag{8}$$

On the other hand, the result is nothing but the First Law of thermodynamics

$$dE \ + \ p\,dV \ = \ T\,dS\ , \tag{9}$$

with $dS = 0$. Here $E = \rho V = \rho a^3 V_0$ is energy, T is temperature and S is entropy of some (fixed) comoving volume V_0. Therefore, Friedmann expansion driven by an ideal fluid is isentropic, $dS = 0$. This is not unexpected, and relaxing the assumption of a perfect fluid will lead to entropy production. However, the dissipation is negligible in cosmological frameworks (except of special moments, like initial matter creation and possible phase transitions, which will be considered separately) and isentropic expansion is a very good approximation. This gives a useful integral of the motion, $S = \text{const}$. On dimensional grounds, $S \propto T^3 a^3 V_0$ and we obtain frequently used relation between the scale factor and temperature in an expanding Universe

$$a \propto \frac{1}{T}\ . \tag{10}$$

To be precise,

$$S = \frac{2\pi^2}{45}\,g_*\,T^3\,a^3 = \text{const}\ , \tag{11}$$

where the factor g_* counts the effective number of relativistic degrees of freedom

$$g_* = \sum_{i=\text{bosons}} g_i + \frac{7}{8}\sum_{j=\text{fermions}} g_j \equiv (g_B + \frac{7}{8} g_F)\ . \tag{12}$$

At any given temperature, only particles with $m \ll T$ should be counted, i. e. g_* is a function of temperature, which is shown in Fig. 1. For a gas of photons, $g_* = 2$. Considering the current epoch of the evolution of the Universe, we have to add the neutrino contribution, which will be discussed later and leads to a different account of effective degrees of freedom at temperatures below e^+e^- annihilation in entropy, g_S, and in energy density, g_ρ. Namely, $g_S(T_0) = 3.909$ and $g_\rho(T_0) = 3.363$. At temperatures above the electroweak scale $g_* \sim 100$ in the Standard Model.

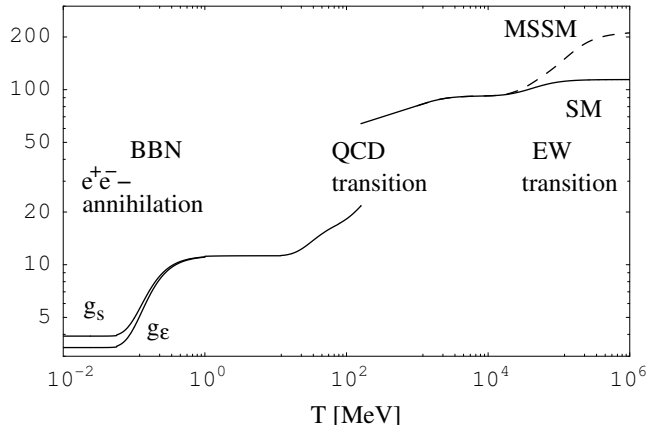


Fig. 1: Number of relativistic degrees of freedom as a function of temperature. From Ref. [12].

Let us give here also other thermodynamical relations, similar to Eq. (11), but for the energy density, ρ, and particle number density, n

$$\rho = \frac{\pi^2}{30}\, g_* T^4 \,, \tag{13}$$

$$n = \frac{\zeta(3)}{\pi^2}\, (g_B + \frac{3}{4} g_F)\, T^3 \,, \tag{14}$$

where $\zeta(3) = 1.202$. These relations are a simple consequence of the integration of Bose–Einstein or Fermi–Dirac distributions

$$\frac{g}{(2\pi)^3} \int \frac{d^3q}{\mathrm{e}^{q/T} \pm 1}\, q^a \,, \tag{15}$$

where q is particle momentum, the plus (minus) sign corresponds to fermions (bosons), and $a = 1$ in calculation of ρ, while $a = 0$ in calculation of n (in the latter case the integral cannot be evaluated in terms of elementary functions and the Riemann ζ-function appears). With the use of Eq. (9), the entropy density, Eq. (11), can be found as $s = 4\rho/3T$, since for relativistic particles $p = \rho/3$ regardless of spin.

2. The Friedmann equation, Eq. (5), can be interpreted within Newtonian mechanics. Indeed, let us first re-arrange it as

$$\frac{1}{2}\,\dot{a}^2 - \frac{4\pi G}{3}\,\rho\, a^2 = -\frac{k}{2} \,. \tag{16}$$

Now, it is easy to see that for $r = a\, r_0$, the Friedmann equation takes the form of energy conservation for test particles bounded in the gravitational potential created by mass $M = \frac{4\pi}{3}\,\rho\, r^3$,

$$\frac{1}{2}\,\dot{r}^2 - \frac{G\,M}{r} = -\frac{k\, r_0^2}{2} \,. \tag{17}$$

We see that the constant k, which determines the sign of spatial curvature in the language of general relativity, also determines the sign of the binding energy

$k = +1$	Binding energy is negative,
	the Universe will recollapse
$k = -1$	Binding energy is positive,
	the Universe will expand for ever

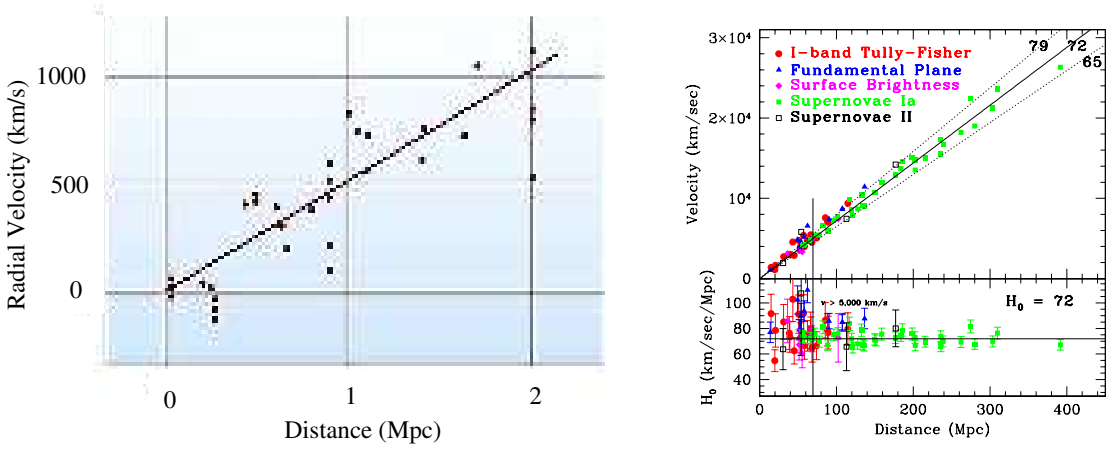


Fig. 2: Hubble Diagram. From Ref. [14].

Therefore, the case of zero spatial curvature, or zero binding energy, $k = 0$, is special and corresponds to fine tuning between initial kinetic and potential energies. Setting $k = 0$ in Eq. (5), this fine tuning can be expressed as $\rho = \rho_c$, where the critical density is defined as

$$\rho_c \equiv \frac{3}{8\pi G} \left(\frac{\dot{a}}{a}\right)^2 . \tag{18}$$

The critical density is proportional to the square of another fundamental parameter,

$$H \equiv \frac{\dot{a}}{a} . \tag{19}$$

The present value of this parameter is called the Hubble constant. It describes the rate of expansion of the Universe, and can be related to observations in the following way. Consider two points with a fixed comoving distance r_0 between them (this means that points do not feel any other forces and do not participate in any other motion beyond general expansion of the Universe). The physical distance between points increases as $r(t) = a(t)r_0$, and we can find the relative velocity as

$$v \equiv \dot{a}r_0 = \frac{\dot{a}}{a} ar_0 = Hr . \tag{20}$$

The relation $v = Hr$ is called the Hubble law. This is shown in Fig. 2. The left panel is original data used by E. Hubble, the right panel presents recent data from Ref. [13]. We will discuss it in more detail later on, especially in relation to observations. But the first thing we may notice is that according to the Hubble law, $v \sim 1$ at $r \sim H^{-1}$. Separations (or wavelengths) of this order are therefore special in cosmology and mean *super-horizon* length scale. At smaller separations, Newtonian gravity should be valid. Einstein's equations tell us that energy conservation in Newtonian mechanics, Eq. (5), and the first law of thermodynamics, Eq. (7), applied to the Universe as a whole, can be extended beyond horizons without any change. The second Friedmann equation, Eq. (6), can be derived as a consequence of these two equations. However, Newton would hardly have done it, even if he had known the first law of thermodynamics. Indeed, in Eq. (6) we recognize Newton's second law, $F = m\dot{v}$ with F being the gravity force, and energy conservation is derived from the equations of motion, not vice versa.

We see that, according to Einstein's theory, the force law is modified. Not only does mass gravitate, but the pressure, too, makes its contribution to the gravitational force. This is a very important modification, since pressure can be negative, leading to anti-gravity and to accelerated expansion. As we

will see, this stage of expansion may have led to the creation of the Universe in our classical understanding. At present, the expansion of the Universe seems to be dominated by anti-gravity as well. This has an interesting history. Newton did not know that one should worry about horizons, but he worried why the Universe does not collapse under the pull of gravity. Einstein was worried about this too. He added (1917) a cosmological constant to the equations of motion, thinking that it will make the Universe static. (As we will see, the cosmological constant corresponds to a vacuum with non-zero energy density and negative pressure, $p = -\rho$.) However, Friedmann had shown in 1922 that the Universe will not be static anyway. After some debate, Einstein admitted his mistake and called the introduction of a cosmological constant "the greatest blunder of my life".

So, why did the Universe not collapse under the pull of gravity? Resolution is in awkward initial conditions called the Big Bang, where velocity in Eq. (17) is highly tuned to potential energy, leading to practically zero spatial curvature and to $\rho = \rho_c$. This implies enormous fine-tuning for the Universe to survive till the present. Such fine-tuning is hard to accept, and a modification of classical cosmology was called for. We will see how modern inflationary cosmology solves the problem of initial conditions. Again, the resolution is in anti-gravity caused by negative pressure.

2.3 Matter content in the Universe

To solve the Friedmann equations, Eqs. (5)–(6), one has to specify the matter content of the Universe and the equation of state for each of the constituents. To fit current observations we need at least four components

- *Radiation* (relativistic degrees of freedom). Today this component consists of the photons and neutrino and gives negligible contribution into total energy density. However, it was a major fraction at early times.
- *Baryonic matter*. Makes up the observable world today.
- *Dark matter*. Was not directly detected yet, but should be there. Constitutes major matter fraction today. Has rather long observational history and can be fitted within frameworks of modern particle physics nicely, at the price of 'moderate' tuning of parameters to provide the required fraction of matter.
- *Dark energy*. Looks like it also should be there. It provides the major fraction of the total energy density today. Was not anticipated and appears as the biggest surprise and challenge for particle physics, though conceptually it can be very simple, being just a 'cosmological constant' or vacuum energy.

Equations of state. Each of these components has a very simple equation of state, parametriz-ed by a single constant w

$$w \equiv \frac{p}{\rho} \ . \tag{21}$$

With a constant w, solutions of the first law of thermodynamics, Eq. (7), are readily found

$$\rho(t) = a(t)^{-3(1+w)} \ \rho_0 \ , \tag{22}$$

where ρ_0 stands for the present-day density. For example, for ordinary forms of matter we have

- Radiation: $w = \frac{1}{3}$ and $\rho = a^{-4} \ \rho_0$. The result can be understood as a simple consequence of entropy conservation, $aT = \text{const}$, since for radiation $\rho \propto T^4$.
- Matter: $w = 0$ and $\rho = a^{-3} \ \rho_0$. The result can be understood as a simple consequence of particle number conservation, $na^3 = \text{const}$ and $\rho = mn$, where m is particle mass.

For hypothetical matter, which may play the role of dark energy, w is negative.

- Network of cosmic strings: $w = -\frac{1}{3}$ and $\rho = a^{-2} \ \rho_0$.

- Network of domain walls: $w = -\frac{2}{3}$ and $\rho = a^{-1}\,\rho_0$.
- Cosmological constant: $w = -1$ and $\rho = \rho_0$. The result can be understood as a consequence of the Lorentz invariance of a vacuum, which restricts T_μ^ν to be proportional to the Kronecker tensor, $T_\mu^\nu = V\,\delta_\mu^\nu$. Comparing with Eq. (4) we find $p = -\rho$, or $w = -1$.

Law of expansion. The Friedmann equation (5) in a spatially flat Universe and with a single matter component, the energy density of which evolves according to Eq. (22), has the solution

$$a = (t/t_0)^{\frac{2}{3(1+w)}} \quad . \tag{23}$$

In particular

- Radiation: $w = \frac{1}{3}$, $a = (t/t_0)^{1/2}$
- Matter: $w = 0$, $a = (t/t_0)^{2/3}$
- Cosmological constant: $w = -1$. This case is special, and $a = \mathrm{e}^{H_0 t}$.

2.4 Cosmological parameters

These are used to parametrize the Fiedmann equation and its solution $a(t)$. Let me first summarize the current knowledge of numerical values of those parameters which were introduced already (Table 1); later in the course we will discuss how these values were deduced.

Table 1: Cosmological parameters

Symbol and definition	Description	Present value
t	Age of the Universe	$t_0 = (13.7 \pm 0.2)$ Gyr
$H = \dot{a}/a$	Hubble constant	$H_0 = 71$ km s^{-1} Mpc^{-1}
$\rho_c = 3H^2/8\pi G$	Critical density	$\rho_c = 10\ h^2$ GeV m^{-3}
$\Omega = \rho/\rho_c$	Omega	$\Omega_0 = 1.02 \pm 0.02$
$\Omega_{\rm CMB} = \rho_{\rm CMB}/\rho_c$	Fraction of CMB photons	$\Omega_{\rm CMB} = 2.4 \cdot 10^{-5} h^{-2}$
$\Omega_b = \rho_b/\rho_c$	Baryonic fraction	$\Omega_b = 0.044 \pm 0.004$
$\Omega_m = \rho_m/\rho_c$	Matter fraction	$\Omega_m = 0.27 \pm 0.04$
$\Omega_\Lambda = \rho_\Lambda/\rho_c$	Dark energy fraction	$\Omega_\Lambda = 0.73 \pm 0.04$

2.5 Cosmography

We can define cosmography (this is my customary definition for these lectures) as a part of cosmology which tries to map observations into reconstruction of the scale factor: The goal is to find and tabulate the function $a(t)$. This is important in many respects: for example, it allows us to determine the matter content in the Universe (assuming the Friedmann equations are correct). One simple and straightforward way of tabulating the function is in determining its coefficients in Taylor expansion. This can be done making a Taylor decomposition around present time, $t = t_0$. The value of the scale factor at any moment of time can be fixed arbitrarily, we can use this freedom to choose $a(t_0) = 1$. The second term in the Taylor decomposition is naturally the value of the Hubble constant, H_0, Eq. (19). It gives us the rate of expansion of the Universe at present and can be measured using the Hubble law, Eq. (20). One can go further in this decomposition and define the second derivative of the Universe at present, the 'deceleration parameter', and so on. We will not do it (at least not at this point), since modern observations probe the whole function $a(t)$. Therefore, let us start with the preparation of the necessary machinery which allows us to deduce $a(t)$ from observations.

Let me stress now that Eq. (20) involves some degree of cheating since it is not a relation between the observables. (However, it is a valid relation for small separations.) To apply the Hubble law to observations, we have to derive its generalization, which would connect quantities we can measure.

Redshift. Looking at distant objects we see only the light they emit. How can physical quantities like distance and velocity be derived? Recall how a police officer determines the speed of a car. A similar principle is used in cosmology to determine the velocity of distant bodies. The shift of emission lines with respect to the frequency measurements by the local observer is related to velocity, and is used as an observable instead of the velocity. Systematic recession of objects, or cosmological expansion, leads to redshift. Note that cosmological redshift is not entirely due to the Doppler effect, but, rather, can be interpreted as a mixture of the Doppler effect and of the gravitational redshift.

Let us relate now the redshift to cosmological expansion, $\dot{a}/a$. To this end, we consider photon trajectories in a cosmological background with metric Eq. (2). The trajectory is given by $ds^2 = 0$. Since the overall scale factor does not change the solutions of $ds^2 = 0$, it is convenient to introduce the conformal time η defined as

$$dt = a(\eta)\, d\eta \; . \tag{24}$$

It is sufficient to consider radial trajectories with the observer at the centre, and I restrict myself to a spatially flat metric $ds^2 = a^2(d\eta^2 - d\chi^2) = 0$, where χ denotes the radial coordinate. The solution of $d\eta^2 - d\chi^2 = 0$ is $\chi = \pm\eta + \text{const}$. Since the comoving distance between source and observer does not change, the conformal time interval between two light pulses is the same at the point of emission and at the point of observation, $\Delta\eta = \text{const}$. Using the definition of conformal time, $d\eta = dt/a$, we find

$$\frac{\Delta t}{a}|_{\text{emission}} = \frac{\Delta t}{a}|_{\text{detection}} \; .$$

Therefore, for a signal frequency we get $\omega_d a_d = \omega_e a_e$. Defining (measurable) redshift as

$$z \equiv \frac{\omega_e - \omega_d}{\omega_d} \tag{25}$$

we obtain

$$1 + z = \frac{a_d}{a_e} \; . \tag{26}$$

It is convenient to normalize the scale factor by the condition $a_d = 1$ at the point of detection, and to consider the scale factor at the point of emission as a function of redshift z. On the basis of this relation, the expansion history of the Universe can be parametrized as

$$a(z) = \frac{1}{1+z} \; . \tag{27}$$

The differential form of Eq. (27) is $da/dz = -a^2$. For future use, let us find now the relation between $d\eta$ and dz

$$d\eta \;=\; \frac{d\eta}{dt}\,\frac{dt}{da}\,\frac{da}{dz}\,dz \;=\; -\,\frac{a}{\dot{a}}\,dz \;=\; -\,\frac{dz}{H(z)} \; . \tag{28}$$

Observing that $d\eta = -d\chi$, we obtain the Hubble low for small separations, $dz = Hd\chi$. At this point, we have succeeded in replacing the velocity by the redshift. Now we aim to relate the distance to some other quantity, directly measurable in cosmology.

Luminosity distance. Looking at distant objects we see only the light they emit. How can physical quantities like distance and velocity be derived? There are several ways to introduce a quantity related to distance: different definitions are not equivalent in curved space–time. A definition based on flux measurements is the appropriate one, if 'standard candles' can be found and used. Detected flux [erg $\text{s}^{-1}\,\text{sm}^{-2}$] is inversely proportional to the distance from a source, $F \propto D^{-2}$. Namely, if L is intrinsic luminosity [erg s^{-1}], we have

$$D_L^2 = \frac{L}{4\pi F} \; . \tag{29}$$

D_L is called the 'luminosity distance'. For this technique to work, one has to find a set of sources with a known or calibrated luminosity. If such sources can be defined, they are called *standard candles.*

To see how the luminosity distance enters the Hubble law, let us consider a space-time with the metric $ds^2 = a^2\ (d\eta^2 - d\chi^2 - \chi^2\ d\Omega)$. Now, go through the following list to find out what happens with the flux emitted into a frequency interval $d\nu$ by a source located at redshif z:

- Surface area at the point of detection is $4\pi\chi^2$. (Recall our choice $a_d = 1$.)
- Energy and arrival rates are redshifted between the points of emission and detection. This reduces the flux by $(1+z)^2$.
- Opposing this tendency, the bandwidth $d\nu$ is reduced by $(1+z)$.
- Photons observed at a frequency ν, were emitted at $(1+z)\,\nu$.

Therefore, the measured spectral flux density is

$$S(\nu) = \frac{L((1+z)\,\nu)}{4\pi\,\chi^2(1+z)}\ . \tag{30}$$

For the bolometric flux (i.e. integrated over ν) we find

$$F = \frac{L}{4\pi\,\chi^2(1+z)^2}\ . \tag{31}$$

Comparing this with the definition, Eq. (29), we find for the luminosity distance

$$D_L = (1+z)\,\chi\ , \tag{32}$$

where χ is the comoving distance between the point of emission and the point of detection

$$\chi(z) = \int_{\eta_e}^{\eta_d} d\eta = \int_0^z \frac{dz'}{H(z')}\ . \tag{33}$$

In the last equality we have used Eq. (28). Therefore, the generalization of the Hubble law, which can be used in observational cosmology, can be written as

$$(1+z)\,\chi(z) = \sqrt{\frac{L}{4\pi F}}\ . \tag{34}$$

Parametrization of H(z). Let us express now the function $H(z)$ in the r.h.s. of Eq. (33) through the cosmological parameters. First, we define the ratio of the total energy density to the critical one

$$\Omega \equiv \frac{\rho_{\rm tot}}{\rho_c}\ . \tag{35}$$

The present day value is referred to as Ω_0. Similarly, for each energy component we denote its *present day* fractional contribution as $\Omega_i \equiv \rho_i/\rho_c$. With these definitions, the Friedmann equation (5) for a spatially flat Universe can be re-written as

$$H^2 = \frac{8\pi G}{3}\sum_i \rho_i\ ,$$

or

$$H^2 = H_0^2\ \sum_i \Omega_i\,(1+z)^{3(1+w_i)}\ . \tag{36}$$

Here I have used

$$\rho_i = \rho_{i,0}\ a^{-3(1+w_i)}\ = \rho_c\,\Omega_i\,(1+z)^{3(1+w_i)}\ ,$$

and expressed the scale factor as a function of redshift according to Eq. (26), and used the definition of the critical density, $H_0^2 = 8\pi G\rho_c/3$.

Parametrization (36) is ready for use in Eq. (33) for the comoving distance. In particular, this finalizes expression Eq. (34) for the luminosity distance.

3 COSMIC MICROWAVE BACKGROUND RADIATION

The Universe is filled with radiation which is left over from the Big Bang. The name for this first light is Cosmic Microwave Background Radiation (CMBR). Measurements of tiny fluctuations (anisotropy) in CMBR temperature give a wealth of cosmological information and became a most powerful probe of cosmology.

This radiation was predicted in 1946 by Georgi Gamov, who estimated its temperature to be ~ 5 K. Gamov was trying to understand the origin of chemical elements and their abundances. Most abundant, after hydrogen, is helium, with its share being $\sim 25\%$. One possibility which Gamov considered was nucleosynthesis of He out of H in stars. Dividing the total integrated luminosity of the stars by the energy released in one reaction, he estimated the number of produced He nuclei. This number was too small in comparison with observations. Gamov assumed then that the oven where the light elements were cooked up was the hot early Universe. He calculated abundances of elements successfully and found that the redshifted relic of thermal radiation left over from this hot early epoch should correspond to ~ 5 K at present. In one stroke, Gamov founded two pillars (out of four) on which modern cosmology rests: CMBR and Big Bang Nucleosynthesis (BBN). The Hot Big Bang was born.

Cosmic microwave background was accidentally discovered by Penzias and Wilson [15] at Bell Labs in 1965 as the excess antenna temperature which, within the limits of their observations, was isotropic, unpolarized, and free from seasonal variations. A possible explanation for the observed excess noise temperature was immediately given by Dicke, Peebles, Roll, and Wilkinson and was published in a companion letter in the same issue [16]. They were preparing dedicated search experiment, but were one month late. Penzias and Wilson measured the excess temperature as $\sim 3.5 \pm 1$ K. It is interesting to note that the first (unrecognized) direct measurements of the CMB radiation was done by T. Shmaonov at Pulkovo in 1955, also as an excess noise while calibrating the RATAN antenna [17]. He published the temperature as (3.7 ± 3.7) K. Prior to this, in 1940, Andrew McKellar [18] had observed the population of excited rotational states of CN molecules in interstellar absorption lines, concluding that it was consistent with being in thermal equilibrium with a temperature of ≈ 2.7 K. Its significance was unappreciated and the result essentially forgotten. Finally, before the discovery, in 1964 Doroshkevich and Novikov in an unnoticed paper emphasized [19] the detectability of a microwave blackbody as a basic test of Gamov's Hot Big Bang model. To me, as a theorist, the detection of CMBR looks nowadays like an easy problem. Indeed, a few per cent of the 'snow' on TV screens is due to CMBR.

The spectrum of CMBR is a perfect black body, with a temperature $T = 2.725 \pm 0.002$ K as measured by modern instruments. This corresponds to 410.4 photons per cubic centimetre or to the flux of 10 trillion photons per second per square centimetre. The temperature is slightly different in different patches of the sky—to 1 part in 100 000. And this is most important: the spectrum of these tiny fluctuations tells us a lot about the fundamental properties of the Universe.

CMBR is the oldest light in the Universe. When registering it, we are looking directly at the deepest past we can, using photons. These photons had travelled the longest distances without being affected by scattering, and geometrically came out almost from the Horizon of the Universe. More precisely, the CMB comes from the surface of the last scattering. We cannot see past this surface. That is because at early times the Universe was ionized and not transparent for radiation. With expansion, it cooled down, and when hydrogen recombined, the universe became transparent. Therefore the CMBR gives us a snapshot of the baby Universe at this time, which is called the time of last scattering. Let us determine when the last scattering occurred in the early Universe.

3.1 Hydrogen recombination

At temperatures greater than a few thousand degrees kelvin, the ionized plasma in the Universe consisted mostly of protons, electrons, and photons, with a small fraction of helium nuclei and a tiny trace of some other light elements. To a good approximation we can consider only the hydrogen. Matter is then

ionized at temperatures higher than the hydrogen ionization energy $E_{\rm ion} = 13.6$ eV. At lower T neutral atoms start to form. The baryonic matter is in thermal equilibrium and the equilibrium fraction of ionized hydrogen can be described by the Saha equation (see Ref. [4] for more details)

$$\frac{n_{\rm e}\, n_{\rm p}}{n_{\rm H}} = \left(\frac{m_{\rm e} T}{2\pi}\right)^{3/2} {\rm e}^{-E_{\rm ion}/T} \ , \tag{37}$$

here $n_{\rm e},\ n_{\rm p}$ and $n_{\rm H}$ are the number densities of electrons, protons, and neutral hydrogen respectively. Plasma is electrically neutral, i.e. $n_{\rm e} = n_{\rm p}$. To find the closed relation for the fraction of ionized atoms, $X \equiv n_{\rm p}/(n_{\rm p} + n_{\rm H}) = n_{\rm p}/n_{\rm B}$, we need the relation between the baryon number density, $n_{\rm B}$, and temperature. This relation can be parameterized with the help of an important cosmological parameter called *baryon asymmetry*

$$\eta = \frac{n_{\rm B}}{n_\gamma} = \frac{n_{\rm p} + n_{\rm H}}{n_\gamma} = (6.1 \pm 0.3) \times 10^{-10} \ , \tag{38}$$

where n_γ is the number density of photons

$$n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3 \ , \tag{39}$$

and $\zeta(3) = 1.202$, see Eq. (14). Baryon asymmetry can be estimated by an order of magnitude by simply counting the number of baryons, $\eta = 2.68 \times 10^{-8} \Omega_b h^2$. This is not the most precise method, though; the value presented in Eq. (38) was obtained from fitting the spectrum of CMBR fluctuations, see below. Nowadays, this is the most precise baryometer. Prior to this, the best estimates were obtained comparing BBN predictions of element abundances to observations. Defining recombination as the temperature when $X = 0.1$, we find $T_{\rm rec} \approx 0.3$eV.

The Universe became transparent for radiation when the mean free path of photons became comparable to the size of the Universe at that time. Photons scatter mainly on electrons and we find that the Universe became transparent when

$$\left(\sigma_{\gamma {\rm e}}\, n_{\rm e}\right)^{-1} \ \sim \ t \ . \tag{40}$$

Here, $\sigma_{\gamma {\rm e}} = 8\pi\alpha^2/3m_{\rm e}^2$ is the Compton cross-section. For the temperature of last scattering we find $T_{\rm ls} \approx 0.26$ eV. Taking the ratio to the current CMBR temeperature we find $z_{\rm ls} \approx 1000$.

3.2 Spectrum is not distorted by red-shift

Prior to recombination photons were in thermal equilibrium. Therefore, at last scattering they have the Planck spectrum

$$n(p) = \frac{1}{\exp(E_{\rm ls}/T_{\rm ls}) - 1} \ .$$

Since then, particle momenta are red-shifted, $p = k/a$. Since photons are massless, $E = p$, their energies are red-shifted at the same rate, $E_0 a_0 = E_{\rm ls} a_{\rm ls}$, and the spectrum becomes

$$n \ = \ \frac{1}{\exp(E_0/a_{\rm ls}T_{\rm ls}) - 1} = \frac{1}{\exp(E_0/T_0) - 1} \ ,$$

where we have used the notation $T_0 \equiv a_{\rm ls} T_{\rm ls}$. Therefore, after decoupling, the shape of the spectrum is not distorted. This statement would not be true for massive particles, $E^2 = (p/a)^2 + m^2$.

Measuring CMBR, we should still see the Planckian spectrum, but with red-shifted temperature. Clearly, this conclusion is true not only for cosmological red-shift, but for the gravitational red-shifts as well. For example, fluctuations in the gravitational potential at the last scattering surface should cause fluctuations in CMBR temperature, but do not distort the spectrum.

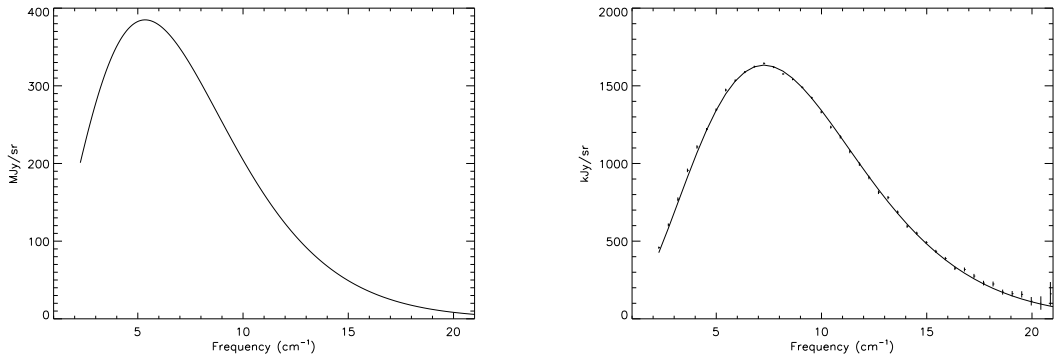


Fig. 3: Left panel: uniform spectrum; error-bars are a small fraction of the line thickness. Right panel: dipole spectrum; vertical lines indicate one σ uncertainties. From Ref. [20].

3.3 Dipole spectrum

Intensity of the CMB radiation is a function of the frequency $2\pi\nu = E$ and the direction on the sky (l, b). As a function of (l, b) it can be decomposed in spherical harmonics. Coefficients will be the functions of ν. The first two terms in this decomposition are

$$S(\nu, l, b) = I_0(\nu) + D(l, b)\, d(\nu) + \dots \ , \tag{41}$$

where $D(l, b) = \cos\theta$, and θ is an angle between observation and the maximum of the dipole $l_0 = 263.85^\circ$, $b_0 = 48.25^\circ$. The dipole is caused by our motion with respect to CMBR (which is composed of the motion of the Sun in the Galaxy and the Galaxy's own motion in the Local Cluster of galaxies). It gives the direction of this motion, l_0, b_0, which roughly coincides with the direction towards Virgo. The dipole induced by the velocity v is $vT\cos\theta$. This gives the magnitude of the Sun's peculiar velocity, (371 ± 1) km s^{-1}.

Let $x \equiv E/T$. The monopole term should have the usual black-body spectrum $I_0(\nu) \propto x^3/(\mathrm{e}^x - 1)$. The dipole spectrum is actually distorted, because the Doppler frequency shift depends upon direction. The dipole spectrum can be found as a term linear in v in the Taylor decomposition of $S(\nu, l, b)$, with the result $d(\nu) \propto x^4\, \mathrm{e}^x/(\mathrm{e}^x - 1)^2$; for a recent discussion see Ref. [21]. Functions $I_0(\nu)$ and $d(\nu)$ are shown in Fig. 3, left and right panels, respectively. Both agree with theoretical expectation.

3.4 Multipoles

Monopole and dipole contributions to CMBR, Eq. (41), can be subtracted. The emission of our Galaxy and various extragalactic sources can be subtracted also. This procedure uses the fact that the relic CMBR signal has a black-body spectrum, which allows us to distinguish it from other forms of radiation: measurements of the intensity at different frequencies allow us to subtract contaminating foregrounds. What remains corresponds to the primordial cosmological pattern of temperature fluctuations, which is shown in Fig. 4. The upper panel presents the results of early COBE experiments [22], the lower panel shows the results of a recent the WMAP experiment [9].

The temperature anisotropy, $T(\mathbf{n})$, as a function of viewing direction vector $\mathbf{n}$, is naturally expanded in a spherical harmonic basis, Y_{lm}

$$T(\mathbf{n}) = \sum_{l,m} a_{lm} Y_{lm}(\mathbf{n}) \ . \tag{42}$$

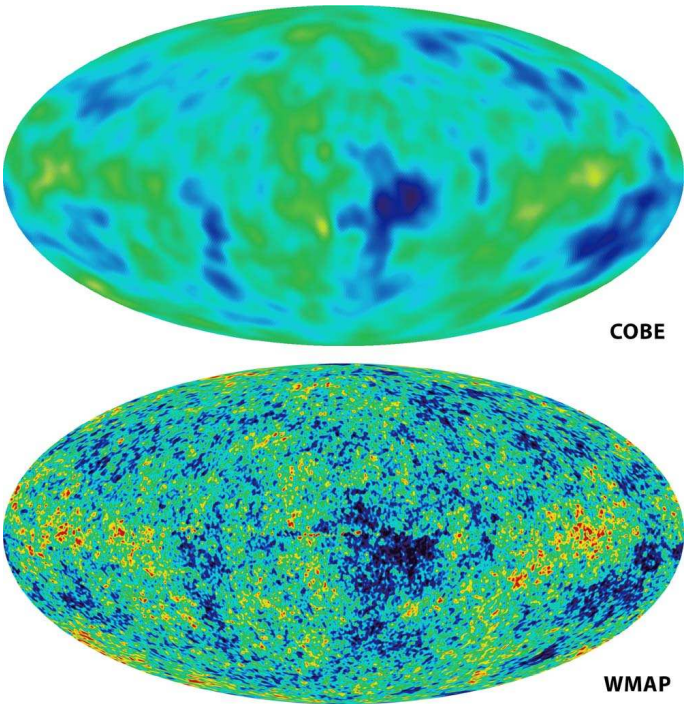


Fig. 4: Pattern of primordial temperature fluctuations in Galactic coordinates, from Ref. [9]. The WMAP map has a resolution 30 times finer than the COBE map.

The coefficients in this decomposition, a_{lm}, define the angular power spectrum, C_l

$$C_l = \frac{1}{2l+1} \sum_m |a_{lm}|^2 \ . \tag{43}$$

The CMBR angular power spectrum as measured by the WMAP experiment is shown in Fig. 5. The harmonic index l is related to the angular scale θ as $l \approx \pi/\theta$, so the first peak, at $l \approx 220$, would correspond to an angular scale of about one degree. Assuming random phases, the r.m.s. temperature fluctuation associated with the angular scale l can be found as

$$\Delta T_l = \sqrt{\frac{C_l \ l(l+1)}{2\pi}} \ . \tag{44}$$

Another representation of temperature fluctuations is given by the angular correlation function, which is related to C_l as

$$C(\theta) = \frac{1}{4\pi} \sum_l (2l+1) C_l \ P_l(cos \ \theta) \ , \tag{45}$$

where P_l is the Legendre polynomial of order l.

3.5 Tool of precision cosmology

The functional form of the CMBR power spectrum is very sensitive to both the various cosmological parameters and to the shape, strength and nature of primordial fluctuations. Measurements of the power spectrum provide us with a wealth of cosmological information at an unprecedented level of precision.

Right after the discovery of CMBR, it was realized that fluctuations in its temperature should have fundamental significance as a reflection of the seed perturbations which grew into galaxies and clusters. In a pure baryonic Universe it was expected that the level of fluctuations should be of the order

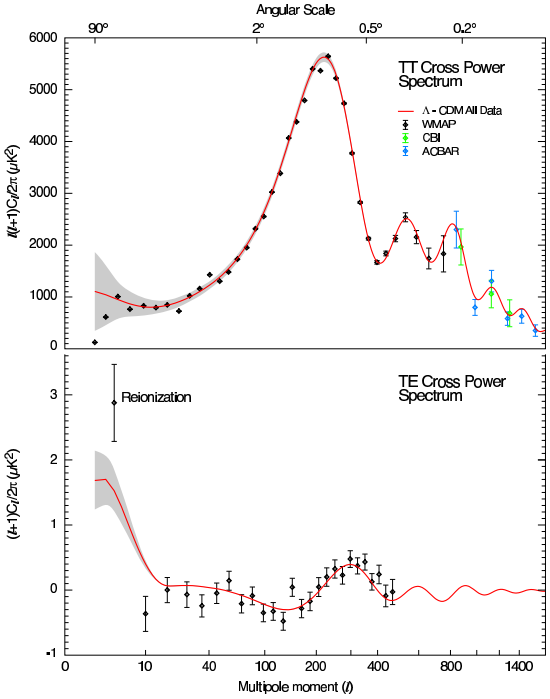


Fig. 5: The angular power spectrum of primordial CMBR temperature fluctuations is shown in the upper panel, from Ref. [9]. Black symbols ($l < 700$) are WMAP measurements, data points at smaller angular scales represent CBI and ACBAR experiments. Lower panel shows the temperature-polarization cross-power spectrum.

$\delta T/T \sim 10^{-2}$–10^{-3}. Measurements of the CMBR anisotropy with ever-increasing accuracy have begun. Once the temperature fluctuations were shown to be less than one part in a thousand, it became clear that baryonic density fluctuations did not have time to evolve into the nonlinear structures visible today. A gravitationally dominant dark matter component was invoked. Eventually, fluctuations were detected at the level of $\delta T/T \sim 10^{-5}$ [23], consistent with structure formation in Cold Dark Matter models with the Harrison–Zel'dovich spectrum of primordial perturbations motivated by cosmological inflation. Already this magnitude of $\delta T/T$ is very restrictive by itself. A partial set of best fit cosmological parameters, as derived from the recent measurments of CMBR anisotropies, is presented in Table 1.

The foundations of the theory of CMBR anisotropy were set out by Sachs and Wolfe [24], Silk [25], Peebles and Yu [26], Syunyaev and Zel'Dovich [27]. The measured spectrum of CMBR power has a characteristic shape of multiple peaks. Positions of these peaks and their relative amplitudes are sensitive to many cosmological parameters in a non-trivial way. Fitting the data to model predictions gives very accurate values for many of these parameters (though there are some degeneracies between deferent sets). Numerical calculations for different models were done already in Ref. [28], and power spectra exhibiting acoustic peaks (similar to those in Fig. 5) were presented. It was realized, in particular, that positions of the peaks are shifted with respect to each other for adiabatic and isentropic primordial fluctuations.

3.6 Acoustic oscillations

Let us give a qualitative picture of why the CMBR power spectrum has a specific shape of a sequence of peaks, and explain how it depends on the values of particular cosmological parameters. Insight, sufficient for the purposes of these lectures, can be gained with the idealization of a perfect radiation fluid. In complete treatment, one has to follow the evolution of coupled radiation and metric fluctuations, i.e. to solve the linearized Einstein equations. However, essential physics of radiation (or matter) fluctuations

can be extracted without going into the tedious algebra of general relativity. It is sufficient to consider the energy–momentum conservation, Eq. (8). To solve for metric perturbations, full treatment based on Einstein equations, Eq. (1), is needed of course. We will not do that here, but simply quote results for the evolution of the gravitational potentials (coincident in some important cases with the solutions for the Newtonian potentials).

Perturbations of the ideal radiation fluid, $p = \rho/3$, can be separated into perturbations of its temperature, velocity and gravitational potential. In the general-relativistic treatment gravitational potential appears as a fractional perturbation of the scale factor in the perturbed metric

$$ds^2 = a^2(\eta)\left[(1+2\Psi)d\eta^2 - (1-2\Phi)dx^i dx_j\right] . \tag{46}$$

Two equations contained in the energy–momentum conservation, $T^{\mu\nu}{}_{;\nu} = 0$ (i.e. temporal $\mu = 0$ and spatial $\mu = i$ parts of this equation), written in metric (46), can be combined to exclude the velocity perturbations. The resulting expression is simple

$$\ddot{\theta}_k + \frac{k^2}{3}\theta_k = -\frac{k^2}{3}\Phi_k + \ddot{\Phi}_k . \tag{47}$$

Note that this equation is the exact result for a pure radiation fluid. Here, θ_k are Fourier amplitudes of $\delta T/T$ with wavenumber k, and Φ_k is a Fourier transform of gravitational potential. Analysis of solutions of the Einstein equations for Φ shows that $\Phi_k = \mathrm{const}$ in two important cases:

1. For superhorizon scales, which are defined as $k\eta \ll 1$.
2. For all scales in the case of matter dominated expansion, $p = 0$.

In these situations the last term in the r.h.s. of Eq. (47), namely, $\ddot{\Phi}_k$, can be neglected. The Einstein equations also restrict the initial conditions for fluctuations. For the adiabatic mode in the limit $k\eta \ll 1$ one finds

$$\delta_i = -2\Phi_i , \tag{48}$$

where $\delta_i \equiv \delta\rho/\rho$. The adiabatic mode is defined as a perturbation in the total energy density. For the one component fluid, which we consider here, only the adiabatic mode can exist. Note that fractional perturbation of the scale factor in metric (46), $a(\eta, \mathbf{x}) = a(\eta) + \delta a(\eta, \mathbf{x}) \equiv a(\eta)(1 - \Phi)$, can be expressed as perturbation of spatial curvature, see Eq. (3). Therefore, adiabatic perturbations are also called curvature perturbations. Let us re-write Eq. (48) for temperature perturbations:

– Radiation domination, $\delta = 4\,\delta T/T$, and we find

$$\theta_i = -\frac{\Phi_i}{2} . \tag{49}$$

– Matter domination, $\delta = 3\,\delta T/T$, and we find

$$\theta_i = -\frac{2\Phi_i}{3} . \tag{50}$$

Recall now that in the limit $k\eta \ll 1$ the gravitational potentail is time-independent, $\Phi = \mathrm{const}$. Therefore, Eq. (47) has to be supplemented by the following initial conditions in the case of the adiabatic mode:

$$\theta_i \neq 0, \qquad \dot{\theta}_i = 0 . \tag{51}$$

Temperature fluctuations on largest scales. Let us consider the modes which had entered the horizon after matter-radiation equality, $k\eta_{\mathrm{eq}} < 1$. For those modes, $\dot{\Phi} = 0$ all the way from initial moments till the present, and the solution of Eq. (47) with adiabatic inital conditions is

$$\theta(\eta) + \Phi = (\theta_i + \Phi)\,\cos\left(\frac{k\eta}{\sqrt{3}}\right) . \tag{52}$$

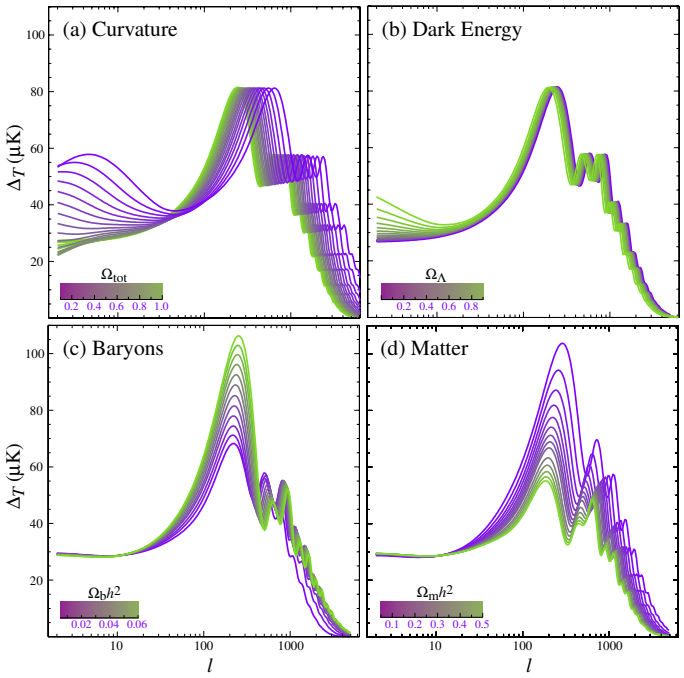


Fig. 6: Sensitivity of the CMBR angular power spectrum to four fundamental cosmological parameters (a) the curvature as quantified by $\Omega_{\rm tot}$ (b) the dark energy as quantified by the cosmological constant Ω_Λ ($w_\Lambda = -1$) (c) the physical baryon density $\Omega_b h^2$ (d) the physical matter density $\Omega_m h^2$, all varied around a fiducial model of $\Omega_{\rm tot} = 1, \Omega_\Lambda = 0.65, \Omega_b h^2 = 0.02, \Omega_m h^2 = 0.147$. From Ref. [29].

As gravity tries to compress the fluid, radiation pressure resists resulting in acoustic oscillations. It is important that oscillations are synchronized. All modes have the same phase regardless of k. This is a consequence of $\dot\theta_i = 0$, which is valid for all k. At the last scattering, the Universe becomes transparent for the radiation and we see a snapshot of these oscillations at $\eta = \eta_{\rm ls}$.

To make its way to the observer, the radiation has to climb out of the gravitational wells, Φ, which are formed at the last scattering surface. Therefore the observed temperature fluctuations are $\theta_{\rm obs} = \theta(\eta_{\rm ls}) + \Phi$, or

$$\theta_{\rm obs} = \frac{1}{3}\Phi_i \cos\left(\frac{k\eta_{\rm ls}}{\sqrt{3}}\right) , \tag{53}$$

where we have used Eq. (50), which relates initial values of θ and Φ. Note that overdense regions correspond to cold spots in the temperature map on the sky, since the gravitational potential is negative. This is the famous Sachs–Wolfe effect [24].

Acoustic peaks in CMBR. Modes caught in the extrema of their oscillation, $k_n\eta_{\rm ls}/\sqrt{3} = n\pi$, will have enhanced fluctuations, yielding a fundamental scale, or frequency, related to the Universe sound horizon, $s_* \equiv \eta_{\rm ls}/\sqrt{3}$. By using a simple geometrical projection, this becomes an angular scale on the observed sky. In a spatially flat Universe, the position of the first peak corresponds to $l_1 \approx 200$, see below. Both minima and maxima of the cosine in Eq. (53) give peaks in the CMBR power spectrum, which follow a harmonic relationship, $k_n = n\pi/s_*$, see Fig. 5.

The amplitudes of the acoustic peaks are recovered correctly after the following effects are taken into account: 1) baryon loading; 2) time-dependence of Φ after horizon crossing in a radiation-dominated Universe; 3) dissipation.

The effect of baryons is exactly the same for the oscillator equation Eq. (52), as if we had increased

the mass of a load connected to a spring and oscillating in a constant gravitational field starting on top of an uncompressed coil at rest. The addition of baryons makes a deeper compressional phase, and therefore increases every other peak in the CMBR power spectrum. (First, third, fifth, . . .) The CMBR power spectrum is a precise baryometer.

Gravitational potentials are not constant, but decay inside the horizon during radiation domination. This decay drives the oscillations: it is timed to leave compressed fluid with no gravitational force to fight when the fluid turns around. Therefore, the amplitudes of the acoustic peaks increase as the cold dark matter fraction decreases, which allows us to measure Ω_m.

Dissipation leads to dumping of higher order peaks in the CMBR power spectrum.

The dependence of the CMBR angular power spectrum on different cosmological parameters is shown in Fig. 6.

The position of the first peak. The Position of the first peak is determined by the angular size of the sound horizon at last scattering. Let us calculate here a similar quantity: the causal horizon (which is larger by a factor of $\sqrt{3}$ in comparison with the sound horizon). The comoving distance travelled by light, $ds^2 = 0$, from the 'Big Bang' to redshift z is determined by a relation similar to Eq. (33), but with different integration limits

$$\eta(z) = \int_z^\infty \frac{dz'}{H(z')} \,, \tag{54}$$

where $H(z)$ is given by Eq. (36). One has to integrate this relation with a complete set of Ω_i, but for simplicity let us consider here the Universe dominated by a single component ρ_j

$$\eta(z) = \frac{(1+z)^{-\gamma_j}}{\gamma_j \, H_0} \,,$$

where $\gamma_j \equiv (1 + 3w_j)/2$. From the last scattering to $z \sim 1$, the Universe was matter dominated. Therefore, the causal horizon in a matter-dominated Universe $(w_j = 0)$

$$\eta(z) = \frac{2}{H_0\sqrt{1+z}}$$

should give a reasonable first approximation.

Consider two light rays registered at $z = 0$ which were separated by a comoving distance $\chi = \eta(z)$ at the moment of emission. Since both propagate in the metric $ds^2 = a^2(d\eta^2 - d\chi^2 - \chi^2 d\theta^2) = 0$, we find for the angular size of horizon at last scattering

$$\theta_h = \frac{\eta(z_{\rm ls})}{\eta(0)} = \frac{1}{\sqrt{1+z_{\rm ls}}} = \sqrt{\frac{T_0}{T_{\rm ls}}} \approx 2^\circ \,. \tag{55}$$

- This gives the position of the first acoustic peak, $l \approx 200$.
- This tells us that there were 10^4 causally disconnected regions at the surface of last scattering.

Horizon problem. Regions separated by more than $> 2^\circ$ have not been in causal contact prior to the last scattering in the standard Friedmann cosmology. The microwave sky should not be homogeneous on scales $> 2^\circ$. Yet, CMB is isotropic to better than 10^{-4} on all scales. Observations tell us that all modes were, indeed, synchronized according to adiabatic initial conditions, Eq. (51), with only small initial perturbations present, $\Phi_i \ll 1$. This constitutes the so-called 'Horizon problem' of standard cosmology. In Section 7 we will see how this problem is solved in the framework of inflationary cosmology.

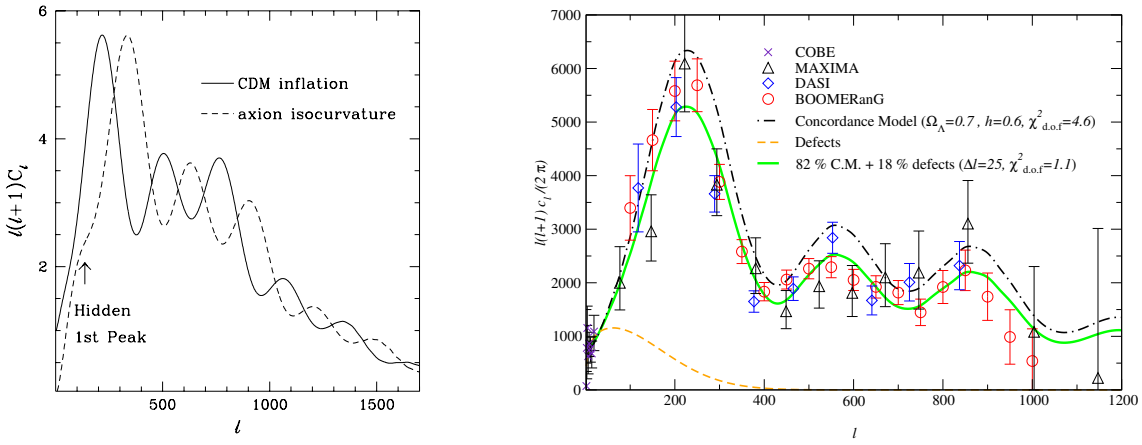


Fig. 7: Left panel: comparison of CMB power spectra in the models with adiabatic and isocurvature initial perturbations, from Ref. [30]. Right panel: adiabatic power spectra in comparison with spectra appearing in models seeded by topological defects, from Ref. [31]. In this panel some older, pre-WMAP, data are also shown.

Non-adiabatic perturbations. For the isocurvature perturbations, instead of Eq. (51), the initial conditions are given by

$$\delta_i = 0, \qquad \dot{\delta}_i \neq 0\ . \tag{56}$$

That is because, in this case, perturbation in total density (and therefore in curvature) are zero initially. As a consequence, in Eq. (52) we will have sine instead of cosine. Acoustic peaks will be shifted by half a period, see Fig. 7. Therefore, isocurvature perturbations are ruled out by modern CMBR experiments.

If density perturbations would be seeded by topological defects (e.g. cosmic strings), both sine and cosine will be present in the solution for temperature fluctuations, Eq. (52). That is because the source for Φ_k is active inside the horizon and phases of θ_k will be random. Acoustic peaks will be absent, see Fig. 7. Structure formation seeded primarily by topological defects is ruled out by modern CMBR experiments.

4 LARGE-SCALE DISTRIBUTION OF GALAXIES

Primordial cosmological fluctuations leave their imprint as CMBR anisotropies (discussed in the previous section), and as density perturbations which give rise to galaxies and clusters of galaxies. CMBR anisotropies are observed on a two-dimensional surface of last scattering, and therefore are measured as a two-dimensional power spectrum. On the other hand, the distribution of galaxies can be measured in three dimensions. (Two angular coordinates of the line of sight to a galaxy and its redshift.) Different physical processes influence the initial perturbations until they are transformed into CMBR fluctuations or fluctuations of the distribution of galaxies. This influence can be encoded as a function of momenta, the transfer function $T(k)$, which simply maps the power spectrum of the initial perturbations into the observed power spectrum, and is a function of cosmological parameters. Therefore, the distribution of galaxies gives complementary information with respect to CMBR anisotropies and helps to break degeneracy between cosmological parameters and the initial spectrum.

This is illustrated in Fig. 8 with CMBR data from WMAP and large-scale structure data from SDSS. The left panel corresponds to 95% constraints in the (Ω_m, h) plane. The shaded dark red region is ruled out by WMAP alone leaving the long banana region. This shows that these two basic cosmological parameters are not well constrained by WMAP alone. The shaded light red region is ruled out when SDSS information is added. The small (shown as white) region remains allowed. Note that the allowed region is in good agreement with a completely independent measurement by the HST key project based

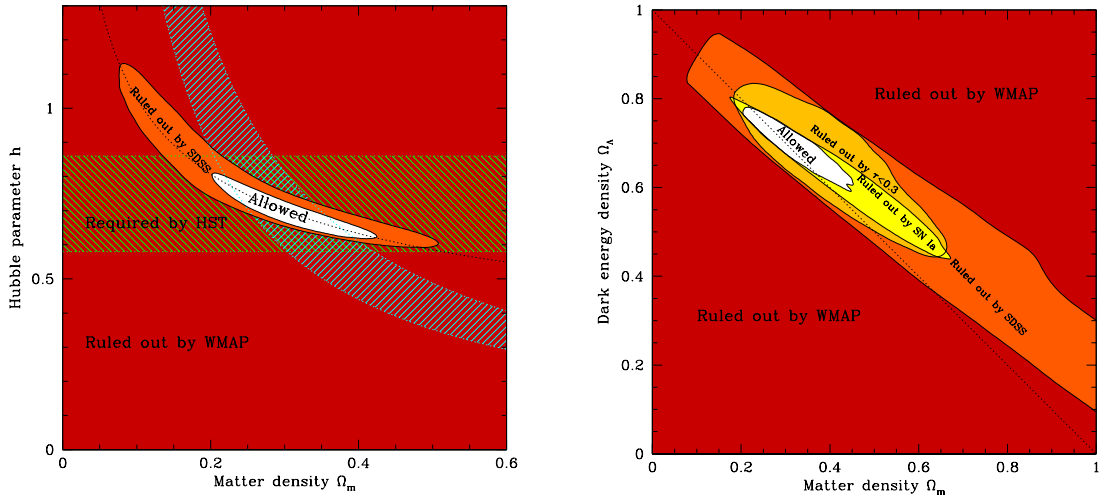


Fig. 8: Combined CMBR and large-scale structure constraints. Left panel: 95% constraints in the (Ω_m, h) plane. Right panel: 95% constraints in the $(\Omega_m, \Omega_\Lambda)$ plane. From Ref. [10].

on entirely different physics. The combined WMAP + SDSS constraint is even tighter than the HST project measurement.

One should bear in mind that there are caveats here. In deriving the WMAP+SDSS constraints which are shown in this figure, it was assumed that the Universe is spatially flat, neutrinos have negligible masses and the primordial spectrum is a pure power law. Without these priors the constraints are less tight.

The constraints in the $(\Omega_m, \Omega_\Lambda)$ plane with the assumption about spatial curvature being relaxed is shown in Fig. 8, right panel. The shaded dark red region is ruled out by WMAP alone, illustrating the well-known geometric degeneracy between models that all have the same acoustic peak locations. The shaded light red region is ruled out when adding SDSS information. Continuing inwards, the next two regions are ruled out allowing the assumption that re-ionization optical depth $\tau < 0.3$ and when supernova SN Ia information is included. $\Omega_\Lambda > 0$ is required with high confidence only when CMBR is combined with galaxy clustering information, or SN Ia information, see the next section.

5 DARK ENERGY

Something which is often called 'dark energy' reveals itself in a variety of cosmological and astrophysical observations. This form of matter gravitates, but does not cluster. Contrary to radiation or dark matter, the dark energy causes the *accelerated* expansion of the Universe. The need for it was hinted at long ago to resolve the conflict between the measured Hubble constant and the lower limits on the age of the Universe. Without the cosmological constant it was also not possible to obtain the correct growth of large-scale structures in the $\Omega = 1$ Universe. Recently, the presence of dark energy was derived from the spectrum CMBR anisotropies and directly detected in the Hubble diagram of high redshift supernovae.

Age of the Universe. If there is no dark energy, the Universe should be matter dominated and should expand according to $a = (t/t_0)^{2/3}$. Differentiating this expansion law we find

$$Ht = 2/3\ . \tag{57}$$

The value of the Hubble constant, as derived by the Hubble Key Project from the Hubble diagram, is 72 ± 8 km s^{-1} Mpc^{-1} [14], see Fig. 2. On the other hand, the lower bound on the age of the Universe

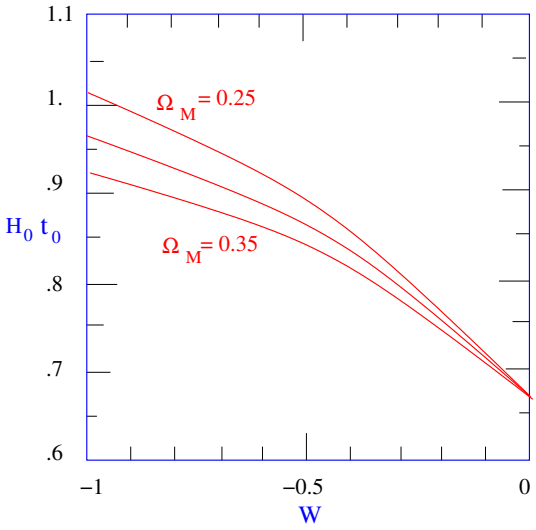


Fig. 9: Graphical representation of Eq. (60) assuming critical density, $\Omega_{\rm tot} = 1$.

can be established estimating the ages of various objects it consists of. For example, the temperature of the coldest white dwarfs in globular clusters yields a cluster age of 12.7 ± 0.7 Gyr [13]. This gives

$$H_0 t_0 > 0.93 \pm 0.12 \,, \tag{58}$$

in clear disagreement with Eq. (57). In other words, the Universe appears much younger than the ages of the oldest objects in it. The critical density Universe, $\Omega = 1$, cannot consist of pressureless matter if measurements of the Hubble constant are correct and Friedmann equations are valid.

The simplest cure (but 'embarrassing' from the point of view of the particle physicist), is to add a cosmological constant, or dark energy. It should be stressed that this minimal modification of Friedmann equations is consistent with all other current cosmological tests and measurements. In the general case, the age of the Universe can be related to the expansion history as

$$t_0 = \int_0^{t_0} dt = \int_0^{t_0} a d\eta = \int_0^{\infty} \frac{dz}{(1+z)H(z)} \,. \tag{59}$$

Here we have used Eqs. (24), (27), (28). For two components, pressureless matter and dark energy with equations of state w, this relation can be written as [see Eq. (36)]:

$$H_0 t_0 = \int_0^{\infty} \frac{dz}{(1+z)^{5/2}\sqrt{\Omega_M + \Omega_{\rm DE}(1+z)^{3w}}} \,. \tag{60}$$

The product $H_0 t_0$ as a function of w, assuming $\Omega_M + \Omega_{\rm DE} = 1$, is shown in Fig. 9. We see that it matches the observational constraints when $w \approx -1$ and $\Omega_M \approx 0.3$.

A discussion of further evidence for dark energy, e.g., related to the problem of the growth of density perturbations, can be found in Ref. [32].

Redshift–luminosity distance relation for Supernovae Ia. For the two-component energy content of the Universe, presureless matter and dark energy, the expression for the luminosity distance, Eq. (33), takes the form

$$D_L = \frac{1+z}{H_0} \int_0^z \frac{dz'}{(1+z')^{3/2}\sqrt{\Omega_M + \Omega_{\rm DE}(1+z')^{3w}}} \,, \tag{61}$$

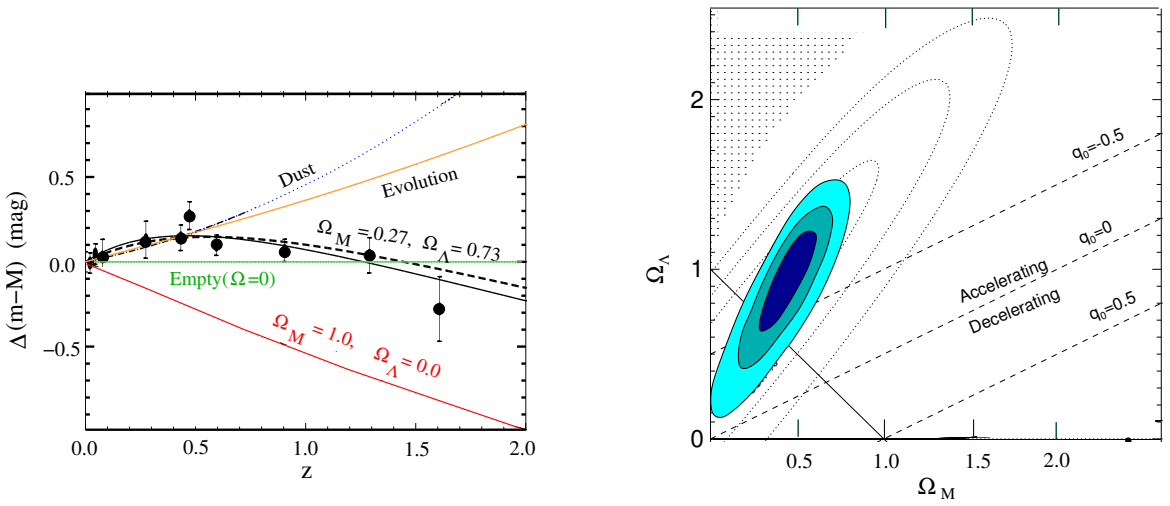


Fig. 10: Left panel: SNe Ia residual Hubble diagram comparing astrophysical models and models for astrophysical dimming. Data (weighted averages in fixed redshift bins) and models are shown relative to an empty Universe model ($\Omega = 0$), adopted from Ref. [11]. Right panel: Joint confidence intervals for (Ω_M,Ω_Λ) from SNe Ia, Ref. [11]. The dotted contours are the results from Ref. [33], illustrating the earlier evidence for $\Omega_\Lambda > 0$. The figure is adopted from Ref. [11].

see Eqs. (32), (33), (36). To use this relation as a cosmological test in conjunction with Eq. (34), one has to find a set of standard candles. This is a big challenge in practice, since we have to find very bright objects which can be seen from far away. At the same time, all of them should have the same luminosity, and we have to be sure that they do not evolve intrinsically. These requirements rule out galaxies and quasars. However, supernovae seem to be suitable. They are bright, as bright as the whole galaxy at the peak of luminosity, and Type Ia supernovae appear to be standard candles. These types of supernovae are thought to be nuclear explosions of white dwarfs in binary systems. The white dwarf, a stellar remnant supported by the degenerate pressure of electrons, accretes matter from a companion and its mass increases toward the Chandrasekhar limit of about $1.4\,M_\odot$. Near this limit, the degenerate electrons become relativistic, which leads to instability and the white dwarf explodes. This physics allows the explosions to be calibrated, since instability occurs under the same conditions.

To proceed, I must remark on the units of flux used by astronomers, which are magnitudes. The system is ancient, and has its origin in the logarithmic response of the human eye. The ratio of the flux of two objects is then given by a difference in magnitudes; i.e.,

$$m_2 - m_1 = -2.5 \log_{10}(F_2/F_1)\ . \tag{62}$$

A smaller magnitude means larger flux.

Figure 10, left panel, shows the corresponding SNe Ia redshift–luminosity distance diagram. Data points correspond to magnitudes of SNe Ia measured at different redshifts. The case of $\Omega_M = 1$ (red curve) is ruled out. The 'concordance model' $\Omega_M = 0.27$ and $\Omega_\Lambda = 0.73$ (black dashed curve) is within 1σ. For a flat geometry prior, best fit corresponds to $\Omega_M = 0.29\pm^{0.05}_{0.03}$ (correspondingly $\Omega_\Lambda = 0.71$). Data are inconsistent with a simple model of evolution of SNe Ia, or dimming due to light absorption by dust as an alternatives to dark energy. The shaded area in Fig. 10, right panel, corresponds to 68%, 95% and 99.7% confidence levels in the (Ω_M, Ω_Λ) plane.

Constraints from CMB. Tight constraints on dark energy, and in a direction in parameter space which is 'orthogonal' to SNe Ie constraints, are obtained from fitting the power spectrum of cosmic microwave

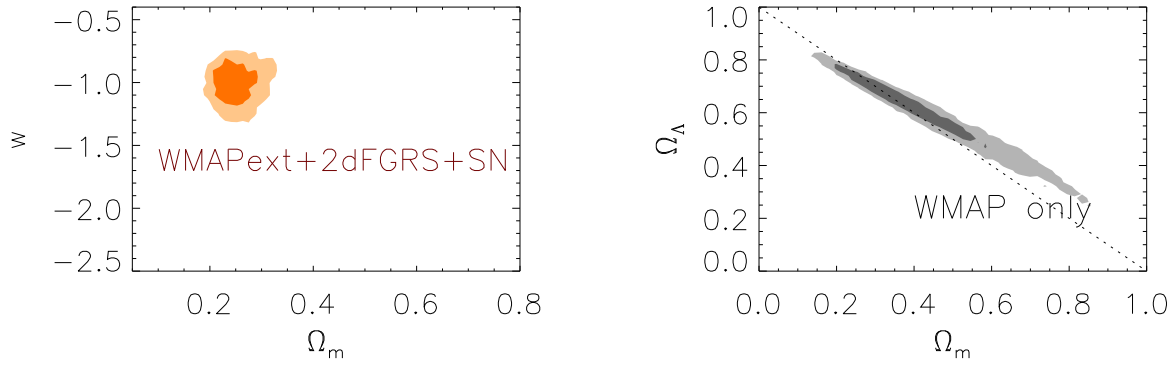


Fig. 11: Left panel: Constraints on the equation of state of dark energy in the (w, Ω_M) plane for a combination of the CMBR + 2dF + SNe Ia data sets. Right panel: Constraints on the geometry of the Universe in the $(\Omega_M, \Omega_\Lambda)$ plane assuming the prior $h > 0.5$. From Ref. [34].

background anisotropies, see the discussion in Section 3 The WMAP data alone rule out the standard $\Omega_M = 1$ CDM model by 7σ if the prior $h > 0.5$ is accepted [9]. The resulting confidence levels in $(\Omega_M, \Omega_\Lambda)$ plane are shown in Fig. 11. While the CMBR data alone are compatible with a wide range of possible properties for the dark energy, the combination of the WMAP data with either the HST key project measurement of H_0, the 2dFGRS measurements of the galaxy power spectrum, or the Type Ia supernova measurements requires that the dark energy be $\Omega_\Lambda = 0.73 \pm 0.04$ of the total density of the Universe, and that the equation of state of the dark energy satisfy $w < -078$ (95% CL) [9].

Constraints from gravitational lensing. Gravitational lensing will be discussed in Section 6. Here we just note that the analysis of strong lensing of sources with known redshift is sensitive to the value of the geometrical cosmological parameters of the Universe. A recent study [35] of the lensing configuration in the cluster Abell 2218 is in agreement with the concordance model. In particular, assuming the flat Universe, it gives for the equation of state of dark energy $w < -0.85$. These constraints are consistent with the current constraints derived from CMB anisotropies or supernovae studies, but they are completely independent tests, providing nearly orthogonal constraints in the $(\Omega_M, \Omega_\Lambda)$ plane, see Fig. 18 in Section 6.

Biggest Blinder – Biggest Surprise. From the point of view of the particle physicist, the cosmological constant just should not be there. Indeed, in quantum field theory, the cosmological constant corresponds to vacuum energy, which is infinite and has to be renormalized,

$$\rho_{\rm vac} = \frac{1}{2\,(2\pi)^3} \int_0^{k_{\rm max}} \omega_k\, k^2\, dk \ . \tag{63}$$

The natural value for the cut-off in this integral is the Planck scale, and then $\rho_{\rm vac} \sim M_{\rm Pl}^4 \approx 10^{74}$ GeV^4. Exact supersymmetry can make this integral vanish. Indeed, in Eq. (63), the contribution of one bosonic degree of freedom is counted. Fermions contribute with an opposite sign, and if there is an equal number of bosons and fermions with equal masses, the vacuum energy will be zero. However, supersymmetry is broken at least at the electroweak scale, and then $\rho_{\rm vac}$ should not be smaller than $\sim M_{\rm W}^4 \approx 10^8$ GeV^4. Before dark energy was detected, it was believed that some yet unknown mechanism reduces the cosmological constant to zero. Zero is a natural number. However, it is hard to understand the smallness of the observed value $\rho_{\rm vac} \approx 10^{-46}$ GeV^4. Moreover, there is another pressing issue of fine tuning: Why does the detected value of $\rho_{\rm vac}$ approximately equal the energy density of

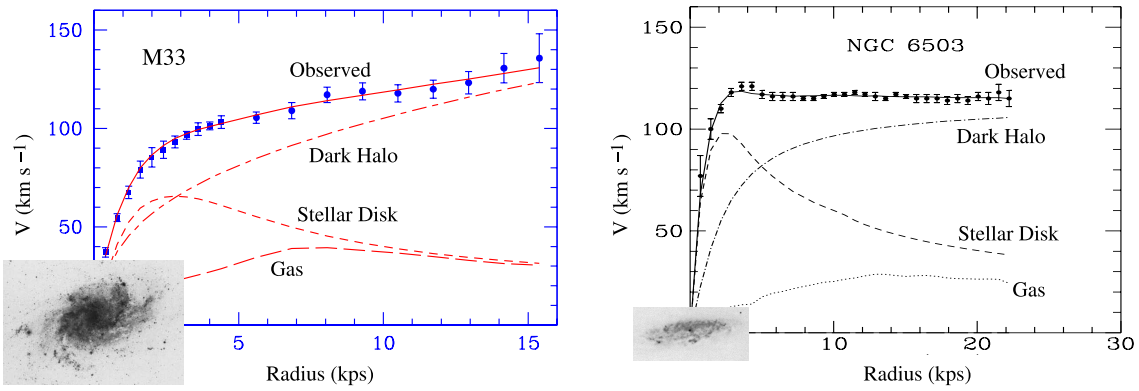


Fig. 12: The rotational curves of two sample galaxies. Left panel: M33, adapted from Ref. [36]. Right panel: NGC6503, adapted from Ref. [37]. I superimposed with the rotational curves the optical images of corresponding galaxies, approximately to scale.

matter at the *present* epoch of cosmological evolution? The ratio of these two contributions scales as a^3 and, say, at recombination the vacuum energy was only 10^{-9} of matter energy. Detection of dark energy not only points to a new physics, but hints that we are missing SOMETHING very fundamental.

6 DARK MATTER

CMBR observations accurately measure the geometry of the Universe, its present expansion rate, its composition, and the nature and spectrum of the primordial fluctuations. Nevertheless, the traditional cosmological tests are still important. In particular, degeneracies between different parameter sets exist, which can produce the same CMBR spectra, and the conclusions drawn do rest upon a number of assumptions. Below we consider cosmological observations that are independent of the CMB and point to the existence of non-baryonic dark matter.

6.1 Dark matter: motivation

The missing mass is seen on all cosmological scales and reveals itself via:

- Flat rotational curves in galaxies.
- Gravitational potential which confines galaxies and hot gas in clusters.
- Gravitational lenses in clusters.
- Gravitational potential which allows structure formation from tiny primeval perturbations.

6.1.1 Dark matter in galaxies

Galactic rotational curves. Consider a test particle which is orbiting a body of mass M at a distance r. Within the frameworks of Newtonian dynamics the velocity of a particle is given by

$$v_{\rm rot} = \sqrt{\frac{G\,M(r)}{r}}\,. \tag{64}$$

Outside of the body, the mass does not depend on distance, and the rotational velocity should obey Kepler's law, $v_{\rm rot} \propto r^{-1/2}$. Planets of the Solar system obey this law. However, this is not the case for stars or gas which are orbiting galaxies. Far away from the visible part of a galaxy, rotational curves

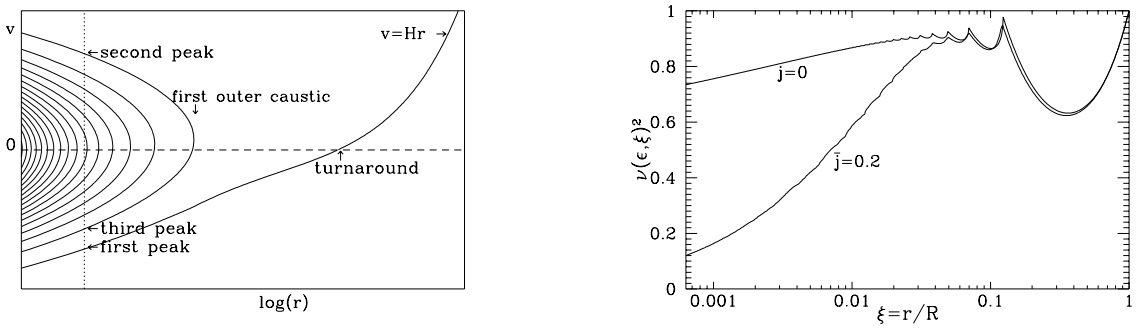


Fig. 13: Left panel: the phase space structure of an infall model. Right panel: rotational curves in an infall models. Two curves which correspond to different angular momenta are shown. From Ref. [40].

are still rising or remain flat. Two examples are shown in Fig. 12. An optical image of the M33 galaxy is superimposed with its rotational curve, approximately to correct scale. The contribution of visible baryons in the form of stars and hot gas can be accounted for, and the expected rotational curve can be constructed. The corresponding contributions are shown in Fig. 12. One can see that the data-points are far above the contribution of visible matter. The contribution of missing dark mass, which should be added to cope with data, is also shown and is indicated as Dark halo. For the rotational velocity to remain flat, the mass in the halo should grow with the radius as $M(r) \propto r$, i.e., the density of dark matter in the halo should decrease as $\rho(r) \propto r^{-2}$.

Halo structure. For direct and indirect dark matter searches it is important to know the phase-space structure of the dark halo as well. With dark matter particles that are interactive, a thermal distribution over velocities would eventually be established. However, in conventional cold dark matter models, particles are non-interacting, except gravitationally. Binary gravitational interactions are negligible for elementary particles, and the resulting phase-space distributions are not unique, even for stationary equilibrium states, and even if flat rotational curves are reproduced.

1. The simplest self-gravitating stationary solution which gives flat rotational curves corresponds to an 'isothermal sphere' with Maxwellian distribution of particles over velocities:

$$n(\vec{r}, \vec{v}) = n(r)\ e^{-v^2/v_0^2}\ . \tag{65}$$

A solution of the equation of hydrostatic equilibrium can be approximated by the density profile

$$\rho(r) = \frac{\rho_0}{(1+x^2)}, \quad \text{where } x \equiv r/r_c\ . \tag{66}$$

It should be stressed that the distribution Eq. (65), in contrast to a distribution in real thermal equilibrium, depends on particle velocities, not on their energies. Such distributions may arise in time-dependent gravitational potential as a result of collisionless relaxation.

2. There exist several density profiles which are empirical fits to numerical simulations, e.g., the Navarro, Frenk and White (NFW) profile [38] and the Moore *et al.* profile [39]

$$\rho(r) = \frac{\rho_0}{x\ (1+x)^2} \qquad \text{NFW}, \tag{67}$$

$$\rho(r) = \frac{\rho_0}{x^{3/2}\ (1+x^{3/2})} \qquad \text{Moore } et\ al. \tag{68}$$

3. In the cold dark matter model, the distribution of particles in the phase space during the initial linear stage prior to structure formation corresponds to a thin hypersurface, $\mathbf{v} = H\mathbf{r}$ (or line in the

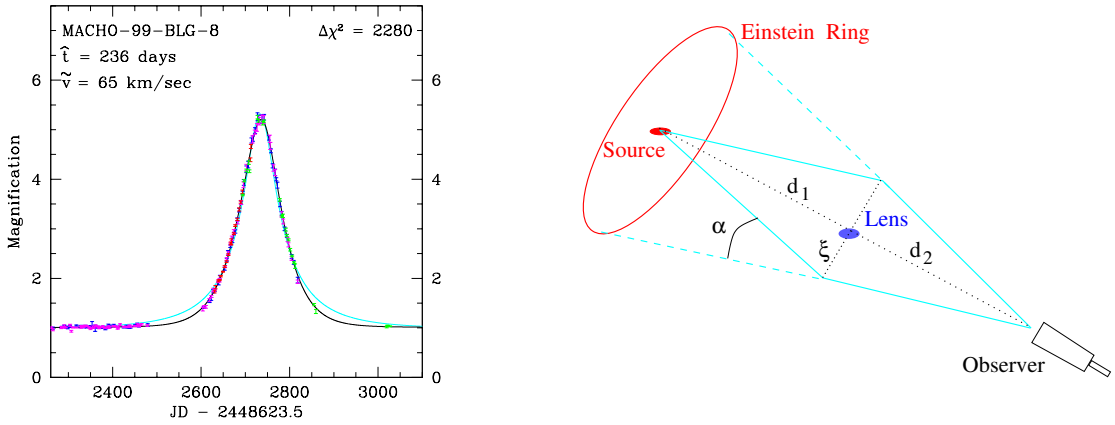


Fig. 14: Left panel: one of the detected microlensing effects, see Ref. [44]. Right panel: schematic view of gravitational lensing by point mass.

Hubble diagram). Since during collisionless evolution the phase-space density conserves, at the non-linear stage the distribution will still be a thin hypersurface. It can be deformed in a complicated way and wrapped around, but it cannot tear apart, intersect its own folds, puff up or dissolve. The corresponding phase-space distribution for the case of spherical symmetry is shown in Fig. 13, left panel. With time, the non-linear structure grows, and the infall of new particles continues. This manifests itself as a growth of turnaround radius (which is a surface where $v = 0$; the turnaround radius of our Galaxy is at 1 Mpc, see Ref. [41]) and as an increasing number of folds inside turnaround. The energy spectrum of dark matter particles at a fixed position will be discrete, see Fig. 13, left panel, where several velocity peaks are indicated at the intersections of the vertical dashed line, $r = \mathrm{const}$, with phase-space sheets. The overall shape of the spectrum also changes compared to an isothermal distribution. This may be important for direct dark matter searches.

The infall model reproduces flat rotational curves, see Fig. 13, right panel. There is one interesting difference, though; rotational curves of the infall model have several small ripples in the region where the curve is flat. These ripples appear near the surfaces where $v = 0$. In principle, they may be detectable [42, 43] and then it will give a clear, unique signature of the presence of dark matter in the galactic halo, (as opposed to models which try to explain apparent violation of Kepler's law by modification of gravity).

The existence of such a folded structure is a topological statement. However, in the inner halo the number of folds is very large, and limited resolution makes the distribution indistinguishable from, say, isothermal. It is not clear at which distances the description of halo in terms of the infall becomes appropriate. But for a sufficiently isolated galaxy, in regions closer to the outer rim of the halo, where the number of folds is still small, signatures of the infall should exist, and they do exist in our Galaxy [41].

Baryonic Halo Dark Matter? No. Already CMBR alone tells us that there should be non-baryonic dark matter, see Table 1. BBN and CMB agree on $\Omega_B = 0.04$, however, the contribution of stars amounts only to $\Omega_{\mathrm{stars}} = 0.005$. There should be dark baryons hiding somewhere. Can it be that the whole, or at least some part, of the halo dark matter is comprised of dark baryons in the form of non-luminous objects? Candidates are Jupiter-like planets, brown dwarfs (which are undersized stars, too light to ignite thermonuclear reactions), or already dead stars (white dwarfs, neutron stars, or even black holes). This class of objects got the acronym MACHO, from MAssive Compact Halo Objects. Special techniques based on gravitational lensing were developed for MACHO searches. These searches were successful, but by now it is clear that MACHOs cannot comprise the whole of dark matter, as their fraction of DM

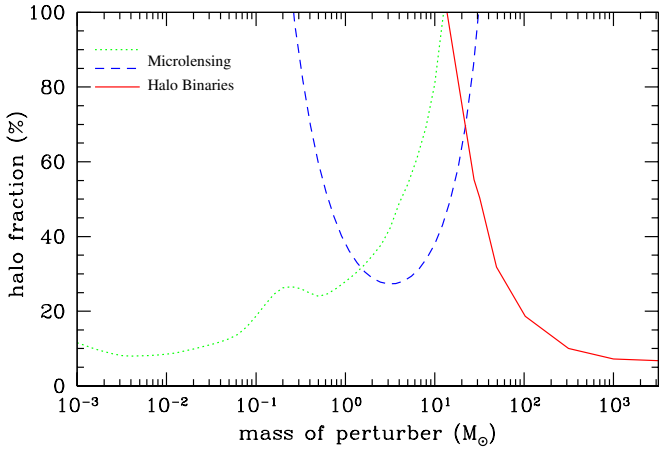


Fig. 15: 95% confidence limits on the MACHO fraction of the standard local halo density. Green and blue lines - results of EROS [45] and MACHO [46] microlensing collaborations. Red line - constraint from the absence of distortion of distribution of binary stars in angular separation [47]. Note that the microlensing exclusion curve extends outside of the plotted range and up to $M \sim 10^{-7} M_\odot$, see Ref. [45].

halo is restricted to be $< 50\%$. Since MACHOs are the only type of dark matter which has been detected, let us consider the issue in some more detail.

Consider the light deflection by a point mass M. If the impact parameter ξ is much larger than the Schwarzschild radius of the lens, $\xi \gg 2GM$, then general relativity predicts that the deflection angle of a light ray, α, is

$$\alpha = \frac{4GM}{\xi} \,. \tag{69}$$

This is twice the value obtained in Newtonian gravity. If the lens happens to be on the line which connects the observer and a source, the image appears as a ring with the radius (Einstein ring radius)

$$r_E^2 = 4\,GM_L\,\frac{d_1 d_2}{d_1 + d_2}\,, \tag{70}$$

see Fig. 14, right panel. If the deflector is displaced from the line of sight by the distance r, then instead of the ring, an odd number of images will appear. If the images cannot be observed separately, because the resolution power of the telescope is not sufficient, then the only effect will be an apparent brightening of the source, an effect known as gravitational microlensing. The amplification factor is

$$A = \frac{2 + u^2}{u\sqrt{4 + u^2}}\,, \tag{71}$$

where $u = r/r_E$. If the lens is moving, the distance r will be changing with time, and the image of the background star will brighten during the closest approach to the line of sight. If the galactic halo is filled with MACHOs, this may happen occasionally for some of the background stars. The typical duration of the light curve is the time it takes a MACHO to cross an Einstein radius, $\Delta t \sim r_E/v$, where $v \sim 10^{-3}$ is typical velocity in the halo. If the deflector mass is $1\,M_\odot$, the average microlensing time will be 3 months, for $10^{-2}\,M_\odot$ it is 9 days, for $10^{-4}\,M_\odot$ it is 1 day, and for $10^{-6}\,M_\odot$ it is 2 hours.

An optical depth for microlensing of the galactic halo is approximately $\tau \sim 10^{-6}$. Thus, if one looks simultaneously at several million stars during an extended period of time, one has a good chance of seeing at least a few of them brightened by a dark halo object. The first microlensing events were reported in 1993. Nowadays, there are more than fifty registered events. One of them is shown in

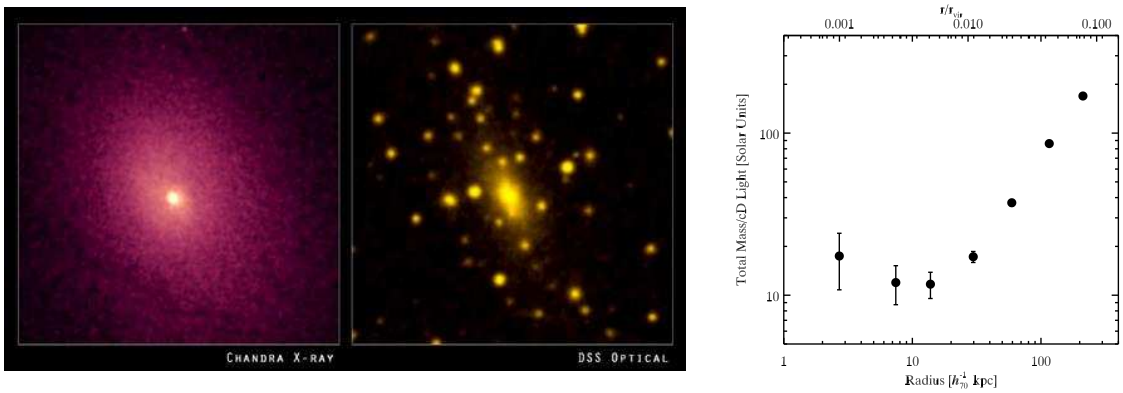


Fig. 16: X-ray (left panel) and optical (middle panel) images of the Abell 2029 cluster of galaxies. Right panel: ratio of total enclosed cluster mass to light in A2029, from Ref. [49].

Fig. 14. However, derived optical depth is not sufficient to account for all dark matter in the Galaxy halo. 95% confidence limits on the MACHO fraction of the standard local halo density is shown in Fig. 15.

Since MACHOs cannot account for the mass of the dark halo, non-baryonic dark matter should be present out there.

6.1.2 Dark matter in clusters of galaxies

In 1933, F. Zwicky [48] deduced the existence of dark matter in the Coma cluster of galaxies. Nowadays, there are several ways to estimate masses of clusters, based on the kinetic motion of member galaxies, on X-ray data, and on gravitational lensing. These methods are different and independent. In the dynamical method, it is assumed that clusters are in virial equilibrium, and the virialized mass is easily computed from the velocity dispersion. In X-ray imaging of hot intracluster gas, mass estimates are obtained assuming hydrostatic equilibrium. Mass estimates based on lensing are free of any such assumptions. All methods give results which are consistent with each other, and tell us that the mass of the luminous matter in clusters is much smaller than the total mass.

Kinetic mass estimates. Those are based on the virial theorem, $\langle E_{\rm pot}\rangle + 2\langle E_{\rm kin}\rangle = 0$. Here $\langle E_{\rm kin}\rangle = N\langle mv^2\rangle/2$ is the averaged kinetic energy of a gravitationally bound object (e.g. cluster of N galaxies) and $\langle E_{\rm pot}\rangle = -N^2\langle m^2\rangle/2\langle r\rangle$ is its averaged potential energy. Measuring the velocity dispersion of galaxies in the clusters and its geometrical size gives an estimate of the total mass, $M \equiv N\langle m\rangle$.

$$M \sim \frac{2\langle r\rangle\langle v^2\rangle}{G} \ . \tag{72}$$

The result can be expressed as mass-to-light ratio, M/L, using the Solar value of this parameter. For the Coma cluster, which consists of about 1000 galaxies, Zwicky [48] found

$$\frac{M}{L} \sim 300\, h\, \frac{M_\odot}{L_\odot} \ . \tag{73}$$

Modern techniques end up with very much the same answer. M/L ratios measured in Solar units in central regions of galaxies range from a few to 10 in spirals and large ellipticals. If clusters are large enough systems for their M/L to be representative of the entire Universe, one finds [50]

$$\Omega_M \approx 0.2 - 0.3 \ . \tag{74}$$

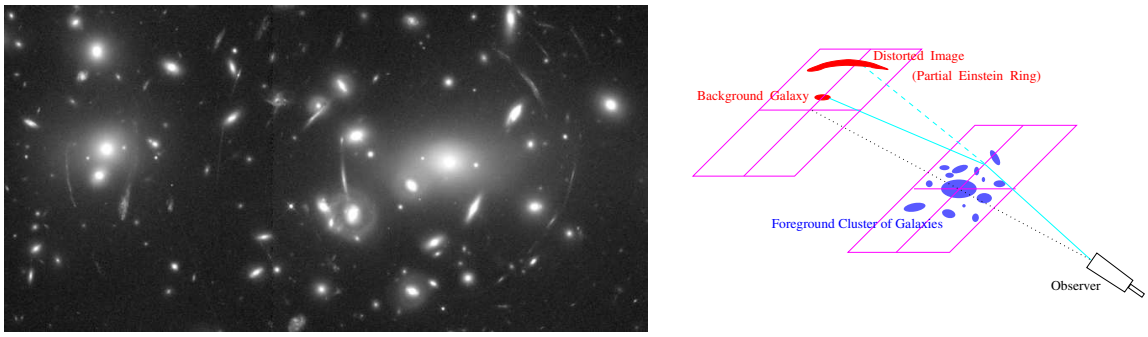


Fig. 17: Left panel: an image of the cluster Abell 2218 taken with the Hubble space telescope (see Ref. [52]). Spectacular arcs resulting from strong lensing of background galaxies are clearly seen.

Mass estimates based on X-rays. Mass is also traced in clusters of galaxies by hot gas which is visible in X-rays. Assume hot gas is in thermal equilibrium in a gravitational well created by a cluster. Then its density distribution $\rho_g(r)$ and pressure $P_g(r)$ satisfy

$$\frac{1}{\rho_g(r)}\frac{dP_g(r)}{dr} = -\frac{GM(\leq r)}{r^2}\,. \tag{75}$$

Observationally, the gas density follows from the X-ray luminosity. Gas temperature can be measured from the shape of the X-ray spectrum. By measuring the temperature profile of a gas, one can reconstruct the gas pressure $P_g(r)$. In this way, the radial run of mass can be deduced.

For example, detailed modelling [49] of Abell 2029, which is shown in Fig. 16, leads to the conclusion that the cluster is dark matter dominated all the way into its core. After subtracting the contributions of stars and hot gas in the mass budget, the density profile of dark matter can be reconstructed. It agrees with the NFW dark matter profile, Eq. (68), $\rho \propto 1/x(1+x^2)$, where $x \equiv r/r_s$ and $r_s = 540$ kpc. The agreement is remarkably good on all scales measured, 3–260 kpc. Baryons contribute $f_b \approx 14\%$ to the total mass of the cluster. Assuming universal baryon mass fraction and Ω_b from big bang nucleosynthesis, this also gives $\Omega_m \approx \Omega_m/f_b \approx 0.29$ for the total mass budget in the Universe, in agreement with other current estimates.

The same methods can be employed for studies of dark matter in large elliptical galaxies. In Ref. [51] the mass profile of the elliptical galaxy NGC 4636, based on the temperature of hot interstellar gas, was obtained for distances from 0.7 to 35 kpc. It was found that the total mass increases as radius to the power 1.2 over this range in radii, attaining a mass-to-light ratio of 40 solar masses per solar visual luminosity at 35 kpc. As much as 80% of the mass within the optical half-light radius is non-luminous in this galaxy.

Gravitational lensing. Gravitational lensing allows direct mass measurement without any assumptions about the dynamical state of the cluster. The method relies on the measurement of the distortions that lensing induces in the images of background galaxies. As photons travel from a background galaxy to the observer, their trajectories are bent by mass distributions, see Fig. 17, right panel. Consider the deflection by a point mass M. For impact parameter ξ which is much larger than the Schwarzschild radius of the lens, $\xi \gg 2GM$, the deflection angle α is given by Eq. (69). If the gravitational field is weak, the deflection angle of an ensemble of point masses will be the vectorial sum of the deflections due to individual lenses.

A reconstruction of lens geometry provides a map of the mass distribution in the deflector. For a review of the method see, for example, Ref [53]. The images of extended sources are deformed by the gravitational field. In some cases, the distortion is strong enough to be recognized as arcs produced by

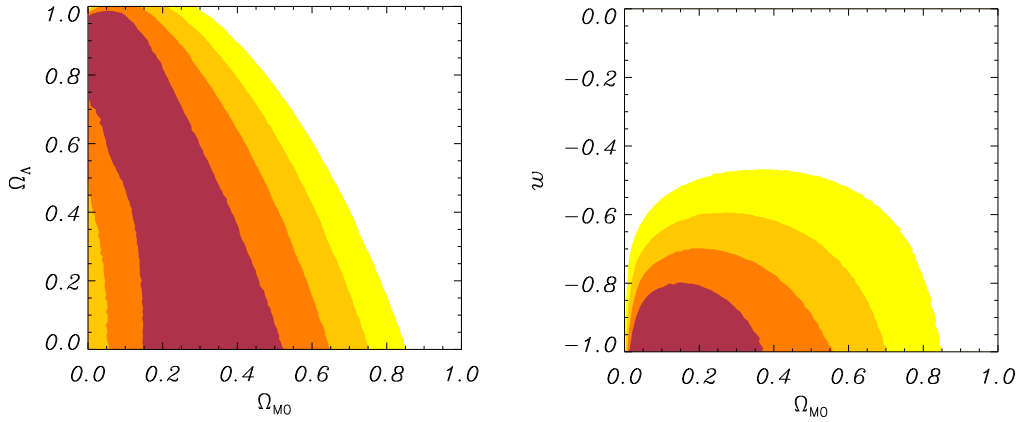


Fig. 18: The constraints on cosmological parameters obtained from the study of the lensing configuration in the Abell 2218 cluster of galaxies [35].

galaxy clusters serving as a lens, see Fig. 17, left panel. For the cluster A 2218, shown in this figure, Squires *et al.* [54] compared the mass profiles derived from weak lensing data and the X-ray emission. The reconstructed mass map qualitatively agrees with the optical and X-ray light distributions. A mass-to-light ratio of $M/L = (440 \pm 80)h$ in solar units was found. Within the error bars the radial mass profile agrees with the mass distribution obtained from the X-ray analysis, with a slight indication that at large radii the lensing mass is larger than the mass inferred from X-rays. The gas to total mass ratio was found to be $M_{\rm gas}/M_{\rm tot} = (0.04 \pm 0.02)\, h^{-3/2}$.

Interestingly, the analysis of multiple images of several sources with known (and significantly different) redshift produced by a cluster lens is sensitive to the value of the geometrical cosmological parameters of the Universe. Study [35] of the lensing configuration in the cluster Abell 2218 gives $0 < \Omega_M < 0.30$ assuming a flat Universe, and $0 < \Omega_M < 0.33$ and $w < -0.85$ assuming a flat Universe with dark energy, see Fig. 18. These constraints are consistent with the current constraints derived with CMB anisotropies or supernovae studies, however, this method is a completely independent test, providing nearly orthogonal constraints in the $(\Omega_M, \Omega_\Lambda)$ plane.

6.2 Structure formation and DM

By now the structure in the Universe (galaxies and clusters) is already formed, the perturbations in matter $\delta\rho/\rho \gtrsim 1$. However, the initial perturbations were small $\delta\rho/\rho \sim 10^{-5}$. Perturbations do not grow in the radiation dominated epoch, they can start growing only during matter domination $\delta\rho/\rho \sim a = 1/z$. Moreover, baryonic plasma is tightly coupled to radiation, therefore perturbations in baryonic matter start to grow only after recombination. For the same reason, initial perturbations in baryons at the time of recombination equal to fluctuations in CMBR. If baryons were to constitute the only matter content, then perturbations in matter at the present time would be equal to

$$\frac{\delta\rho}{\rho}|_{\rm today} = z_{\rm rec}\, \frac{\delta\rho}{\rho}|_{\rm rec} \sim 10^{-2}\ , \tag{76}$$

where $z_{\rm rec} \approx 1100$ is the redshift of recombination. This is one of the strongest and simplest arguments in favour of non-baryonic dark matter. Structure has had time to develop only because perturbations in non-baryonic dark matter have started their growth prior to recombination. Baryonic matter then 'catches up' simply by falling into already existing gravitational wells. If one aims to explain things by modification of gravity, one has to explain not only the flat rotational curves in galaxies and the presence of dark

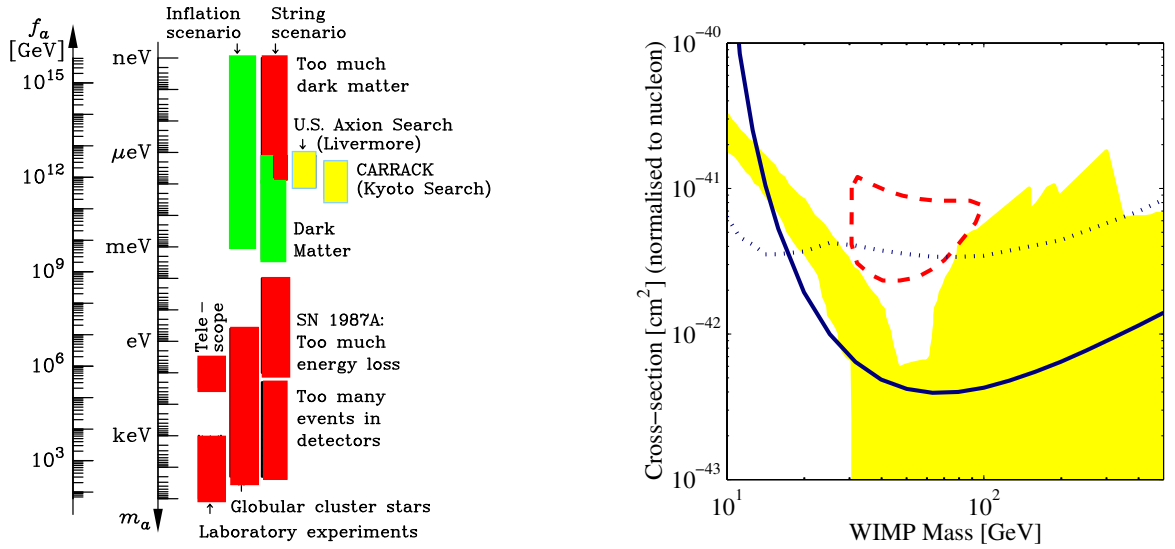


Fig. 19: Left panel: The constraints on axion parameters. Red blocks—various cosmological and astrophysical constraints; yellow blocks—exclusion regions obtained in dedicated dark matter search experiments; green blocks—the allowed regions in two cosmological scenarios. Right panel: Solid curve—the exclusion limits on the coherent WIMP-nucleon scalar cross-section obtained by CDMS collaboration in the direct dark matter search experiment; the parameter space above the curve is excluded at the 90% C.L. These limits are inconsistent with the DAMA 3σ signal region [73] (closed dashed contour) if it is due to scalar coherent WIMP interactions. Also shown are limits from CDMS at SUF (dotted line). The typical predictions of supersymmetric models are shown in yellow. Adapted from Ref. [74].

matter in galaxy clusters, but also to provide the accelerated growth of structure from recombination till the present, in a consistent way.

6.3 Non-baryonic dark matter candidates

There is no lack of dark matter candidates in particle physics models. Some of them appear naturally, and were motivated by some other reasoning, not related to the dark matter problem. They are the leading candidates and are listed below.

- Axion (mass $m \sim 10^{-5}$ eV). Appeared [55, 56] as a by-product of a suggested solution of the strong CP problem via a global $U(1)$ Peccei–Quinn symmetry [57]. The axion picks up a small mass in a way similar to the pion when chiral symmetry is broken. The parameters of these two particles are related, in particular $m_a \sim m_\pi f_\pi / f_a$, where f_a, is the axion decay constant, and determines also the strength of the axion coupling to other particles. There are tight astrophysical and cosmological bounds on f_a which leave only a narrow window, 10^{10} GeV $\lesssim f_a \lesssim 10^{12}$ GeV, see Fig. 19, the left panel. For a review of axion physics and searches see, e.g., Refs. [58, 59].
- Neutrino ($m \sim 0.1$ eV). The only dark matter candidate which is known to exist. For this reason we discuss the neutrino in some more detail below. For a review of neutrino cosmology see, for example, Ref. [60], and for the neutrino astrophysics see Ref. [61]. While the neutrino is cosmologically important, it cannot resolve the dark matter problem.
- Mirror matter ($M \sim 1$ GeV). Does not belong to the list of the most popular candidates, but is attractive as an example of a model [62–64] where the approximate equality of baryonic and non-baryonic contributions in the energy balance of the Universe is attempted to be explained naturally, and not as a result of fine-tuning of model parameters.

– WIMP ($m \sim 100$ GeV). The most popular candidate, a natural outcome of supersymmetry. The lightest supersymmetric particle (or LSP) is naturally stable and would have interesting cosmological abundance. Known also under the names neutralino (dark matter has to be colour and electrically neutral) and WIMP (from Weakly Interacting Massive Particle). For recent reviews see, for example, Refs. [65, 66]. The current status of direct and indirect WIMP searches is reviewed in Ref. [67]. The new limits obtained by the CDMS Collaboration and not reflected in cited reviews are shown in Fig. 19 (right panel).
– WIMPZILLA ($m \sim 10^{13}$ GeV). The newcomer, was initially motivated as a solution of the Greizen–Zatsepin–Kuzmin puzzle of ultra-high-energy cosmic rays [68, 69]. Its popularity was boosted by the observation that cosmologically interesting abundance is created naturally, just as a sole consequence of expansion of the Universe [70, 71].

Non-baryonic dark matter model building and searches make an extensive subject on their own. There are many dedicated excellent reviews, I cannot list them all, see, for example, an earlier one [72] and the latest [65–67] ones. For this reason, and because of space limitations, I will not describe non-baryonic dark matter in all its variety, instead I'll spend some time on simple and universal relations.

Cosmological density of neutrino. Here we calculate the abundance of particles which were once in thermal equilibrium with the rest of cosmological plasma. Let us first consider the case of the neutrino.

Comparing the weak interaction rate, $\Gamma \sim T^5/M_W^4$, to the expansion rate, $H \sim T^2/M_{\rm Pl}$, one finds that neutrinos are in thermal equilibrium at temperatures $T \gg 1$ MeV and decouple from the rest of plasma at lower temperatures. (One can do this in full detail, see Ref. [60].) Therefore, Standard Model neutrinos, which have small masses, decouple when they are still relativistic. The number density of neutrino at this time is given by Eq. (14). Below this temperature, neutrinos are no longer in thermal equilibrium with the rest of the plasma, and their temperature simply decreases as $T \propto 1/a$. However, the cosmological background of photons is heated up by the e^+e^- annihilations. Let us find a relation between T_ν and T_γ, which will also give the relation between n_ν and n_γ.

The annihilation reaction rate is much faster than the expansion of the universe, therefore this process is adiabatic, and entropy in comoving volume is conserved, $g_*\, T^3 = \text{const}$, see Eq.(11). Before annihilation $g_* = g_\gamma + g_e \cdot (7/8) = 2 + 4 \cdot (7/8) = 11/2$. After annihilation $g_* = g_\gamma = 2$. Since before annihilation $T_\nu = T_\gamma$, we find $T_\nu = (4/11)^{1/3}\, T_\gamma$, and for one neutrino flavour we have

$$n_\nu = \frac{3}{11}\, n_\gamma \ . \tag{77}$$

Here we used Eq. (14) and $g_\gamma = g_\nu = 2$, since right-handed neutrinos do not contribute (are not excited) even if the neutrino has a small Dirac mass, see Ref. [60]. As a consequence of e^+e^- annihilation, the neutrino temperature is lower. With $T_\gamma = 2.728$ K we find $T_\nu = 1.947$ K and $n_\nu = 115\ \text{cm}^{-3}$.

At temperatures larger than neutrinos mass, $T_\nu \gg m_\nu$, in the Standard Model, assuming no chemical potential, we find

$$\Omega_\nu = 3 \left(\frac{7}{8}\right) \left(\frac{4}{11}\right)^{4/3} \Omega_\gamma \approx 0.68\, \Omega_\gamma \ , \tag{78}$$

where the factor of 3 corresponds to the number of neutrino flavours. This result allows us to find the epoch of equal matter and radiation densities

$$1 + z_{\rm eq} = \frac{\Omega_M}{\Omega_\gamma + \Omega_\nu} \approx 3200 \ . \tag{79}$$

Assume that by now neutrinos became non-relativistic, i.e. their masses are larger than the present temperature. In this case, neutrino energy density is given by $\rho_\nu = \sum_i m_{\nu i}\, n_{\nu i}$. Since it has to be smaller than $\Omega_m\, \rho_c$, we have the constraint [75]

$$\sum_i m_{\nu i} < 94\, \Omega_m h^2 \text{ eV} = 12 \text{ eV} . \tag{80}$$

For dark matter particles to boost the structure formation, their typical velocities squared at the time of recombination should be smaller than the depth of typical gravitational wells, $v^2 \ll 10^{-5}$. In other words, dark matter should be *cold*. This is not the case for particles as light as those which satisfy the bound Eq. (80). Neutrinos can make up dark matter, but it will be *hot* dark matter.

Neutrino mass is pinned down. Free streaming of relativistic neutrinos suppresses the growth of fluctuations until ν becomes nonrelativistic at $z \sim m_j/3T_0 \sim 1000\ (m_j/\text{eV})$. This effect of free-streaming is not seen in the data and therefore only small corrections due to light neutrinos are allowed in the standard CDM picture. Combined CMBR and LSS analysis yields the constraint [9]

$$\Omega_\nu h^2 = \frac{\sum_i m_i}{93.5 \text{ eV}} < 0.0076 , \tag{81}$$

which translates into the upper bound

$$\sum_i m_i < 0.7 \text{ eV} \qquad (95\%\,\text{CL}) . \tag{82}$$

On the other hand, atmospheric neutrino oscillations provide a lower bound on the heaviest neutrino mass, since $\sqrt{\delta m^2_{\text{atm}}} \sim 0.03$ eV. Combining these two limits

$$0.03 \text{ eV} \le m_{\text{heaviest}} \le 0.24 \text{ eV} \tag{83}$$

we see that the heaviest neutrino mass is now known to within an order of magnitude [76].

Can neutrinos make up a galaxy halo? By $z \sim 1$ the neutrino quanta satisfying the mass bound Eq. (83) became sufficiently non-relativistic to make their way into the gravitational wells. The question arises, can neutrinos at least make up the dark haloes and be responsible for flat rotational curves? The answer to this question is: No. To prove it, let us assume that neutrinos do build up a dark matter halo with a flat rotational curve

$$\rho_{\text{DM}} = \frac{M_{Pl}^2 v_{\text{rot}}^2}{r^2} . \tag{84}$$

We can express the energy density $\rho_{\text{DM}} = \rho_\nu$ through the integral of phase space density over the momenta

$$\rho_\nu = \frac{m_\nu}{(2\pi)^3} \int d^3k\ n(k,r) . \tag{85}$$

But for fermions, the phase-space density, $n(k,r) = n(E)$, should obey the Pauli exclusion principle, $n(E) < 1$. Combining Eqs. (84) and (85), we find $m^4 v^3 \sim M_{Pl}^2 v^2/r^2$, or

$$m_\nu > 120 \text{ eV} \left(\frac{100 \text{ km s}^{-1}}{v_{\text{rot}}}\right)^{1/4} \left(\frac{1 \text{ kpc}}{r_c}\right)^{1/2} . \tag{86}$$

For dwarf galaxies this constraint (the Tremaine–Gunn limit [77]) reads $m_\nu > 500$ eV, and we arrive at a contradiction with Eq. (80), which becomes even stronger when compared with Eq. (83).

Cosmological density of other thermal relics. Assume now that some weakly interacting particle has a mass larger than 1 MeV and decouples when it is non-relativistic. The equilibrium number density will be Boltzmann-suppressed in this case by the exponent $\exp(-m/T)$. The weak interaction cross-section implies $\sigma \sim m^2/M_W^4$, if $m \ll M_W$. Repeating calculations for abundances, one finds that in this case $\Omega_m h^2 \approx 3\ (1\ \mathrm{GeV}/m)^2$, i.e. a correct cosmological abundance of dark matter would be achieved for $m \approx 5$ GeV.

On the other hand, if $m \gg M_W$, the annihilation cross-section becomes $\sigma \sim 1/m^2$ and one finds $\Omega_m h^2 \approx (m/1\ \mathrm{TeV})^2$, i.e. the correct cosmological abundance of dark matter is achieved for $m \approx 300$ GeV. Using field-theoretical unitarity and the observed density of the Universe, it can be shown that a stable elementary particle which was once in thermal equilibrium cannot have a mass greater than 340 TeV [78].

Cosmological density of non-thermal relics. The mass of non-thermal relics can be much larger than $O(10^2)$ TeV without violating the unitarity bound; it can also be much smaller than than $O(1)$ GeV and dark matter still will be cold, as required by observations.

1. Axions. Very light scalar particles, like axions, are created in a state of coherent oscillations. This can be viewed also as a Bose condensate. To illustrate the general idea, let us consider a scalar field with potential $V(\phi) = m^2\phi^2/2$. The equations of motion for the Fourier modes with a momentum k in an expanding Universe are

$$\ddot{\phi}_k + 3H\dot{\phi}_k + (k^2 + m^2)\phi_k = 0\ . \tag{87}$$

Since the term $\propto H$ can be understood as friction, the amplitude of those modes for which $9H^2 \gg (k^2 + m^2)$ (almost) does not change with time. The oscillations of modes with a given k commence when H becomes sufficiently small, $9H^2 \ll (k^2 + m^2)$. Oscillating modes behave like particles, and their amplitude decreases with expansion. Since modes with the largest k start oscillations first, they will have the smallest amplitude and the field becomes homogeneous on a current horizon scale. Modes with all k will already be oscillating when $3H \approx m$, and will behave like cold dark matter after that. Note that the field will be homogeneous on the horizon scale at this time, but may be inhomogeneous on larger scales. This may lead to formation of dense clumps, 'axion mini-clusters' [79, 80] of the mass $M \sim 10^{-12}\, M_\odot$ [81]. Note also that in the case of axions, one has to take into account the dependence of m on temperature T. Solving $3H(T) = m(T)$ one finds $\Omega_{\mathrm{axion}} \sim 1$ when $f_a \sim 10^{12}$GeV [82, 83].

2. Superheavy dark matter. Non-conformal quantum fields cannot be kept in a vacuum in an expanding Universe. This can be understood on the example of a scalar field, Eq. (87). In conformal time, Eq. (24), and for the rescaled field, $u_k \equiv \phi_k\, a$, the mode equations take the form of an oscillator equation

$$\ddot{u}_k + \omega_k^2 u_k = 0\ , \tag{88}$$

with time-dependent frequency

$$\omega_k^2 = k^2 + a^2 m^2 - \frac{\ddot{a}}{a}(1 - 6\xi)\ . \tag{89}$$

The constant ξ describes the coupling to the scalar curvature, the corresponding term in the Lagrangian is $\xi R\phi$. The case of $\xi = 0$ corresponds to minimal coupling [Eq. (87) was written for this case], while $\xi = 1/6$ is the case of conformal coupling. Equations for massless, conformally coupled quanta are reduced to the equation of motion in Minkowski space–time. Particle creation does not occur in this case. For massive particles, conformal invariance is broken and particles are created regardless of the value of ξ. Let us consider the case of $\xi = 1/6$ (the general situation is considered in Ref. [84]). It is the particle mass which couples the system to the background expansion and serves as the source of particle creation in this case. Therefore, we expect that the number of created particles in a comoving volume is $\propto m^3$ and the effect is strongest for the heaviest particles. In fact, stable particles with $m > 10^9$ GeV would overclose the Universe in the standard 'pre-inflationary' Friedmann model [71]. Inflation cuts the

particle production and $\Omega_{\rm SH} \sim 1$ if $m > 10^{13}$ GeV and the reheating temperature is $T \sim 10^9$ GeV, which is the value of reheating temperature compatible with supergravity models [70, 71, 84].

7 BASICS OF INFLATION

In frameworks of 'classical' cosmology and assuming no fine-tuning, one concludes that a typical universe should have had Planckian size, live Planckian time and contain just a few particles. This conclusion is based on the observation that Fridmann equations contain a single dimension-full parameter $M_{\rm Pl} \sim 10^{19}$ GeV, while dimensionless parameters naturally are expected to be of order unity. Yet, the observable Universe contains 10^{90} particles in it and had survived 10^{65} Planckian times. Where does it all came from? In other words, why is the Universe so big, flat ($\Omega_0 \approx 1$) and old ($t > 10^{10}$ years), homogeneous and isotropic ($\delta T/T \sim 10^{-5}$), why does it contain so much entropy ($S > 10^{90}$) and does not contain unwanted relics like magnetic monopoles? These puzzles of classical cosmology were solved with the invention of Inflation [85]– [89]. All these questions are related to the initial conditions and one can simply postulate them. The beauty of Inflation is that it prepares these unnatural initial conditions of the Big Bang, while the pre-existing state (which can be arbitrary to a large extent) is forgotten. Inflationary theory came with unplanned bonuses. Not only does the Universe become clean and homogeneous during inflation, but also the tiny perturbations necessary for the genesis of galaxies are created with the correct magnitude and spectrum. Below we consider the basics of inflationary cosmology.

7.1 Big Bang puzzles and inflationary solutions

Horizon problem and the solution. The size of a causally connected region (horizon) scales in proportion to time, $R_{\rm H} \propto t$. On the other hand, the physical size of a given patch grows in proportion to the scale factor, $R_{\rm P} \propto a(t) \propto t^\gamma$. The exponent γ depends upon the equation of state, $\gamma = 1/2$ for radiation and $\gamma = 2/3$ for matter dominated expansion. In any case, for the 'classical' Friedmann Universe, $\gamma < 1$ and the horizon expands faster than the volume. Take the largest visible patch today. It follows that in the past its physical size should have been larger than the size of the horizon at the time (since they are equal today) and therefore this patch should have contained many casually disconnected regions. For example, as we have found in Section 3, the angular size of the horizon at the moment of last scattering is $\approx 2^\circ$, see Eq. (55), which tells us that we observe 10^4 causally disconnected regions at the surface of last scattering. The question arises, Why is the Universe so homogeneous at large scales?

This problem can be solved if during some period of time the volume expands faster than the horizon. During such a period, the whole visible Universe can be inflated from one ('small') causally connected region. Clearly, this happens if $\gamma > 1$, which means $\ddot{a} > 0$. Either of these two conditions can be used as a definition of an inflationary regime. Using the Friedmann equation (6) we find that the inflationary stage is realized when $p < -\rho/3$. In particular, if $p = -\rho$ the energy density remains constant during expansion in accord with the first law of thermodynamics, Eq. (7), and the physical volume expands exponentially after a few Hubble expansion times, $a(t) = {\rm e}^{Ht}$, see Eq. (5).

Curvature problem and the solution. The Friedmann equation (5) can be re-written as

$$k = a^2 \left(\frac{8\pi G}{3} \rho - H^2 \right) = a^2 H^2 \ (\Omega - 1) = \dot{a}^2 \ (\Omega - 1) = {\rm const}. \tag{90}$$

Here we immediately see the problem: during matter or radiation dominated stages, $\dot{a}^2$ decreases (in general, this happens for any expansion stage with $\ddot{a} < 0$), therefore Ω is driven away from unity. To observe $\Omega \sim 1$ today, the observer has to live in a universe with extreme initial fine-tuning, say at the epoch of nucleosynthesis, when the temperature was $T \sim 1$ MeV, one finds $|\Omega(t_{\rm NS}) - 1| < 10^{-15}$, and even stronger tuning is required at earlier epochs. A possible solution is obvious: accelerated expansion $\ddot{a} > 0$ increases $\dot{a}$ and therefore drives $\Omega(t)$ to unity. A robust, crucial and testable prediction of inflationary cosmology is a flat Universe, $\Omega = 1$.

The problem of entropy. As we know already, the energy of a vacuum, $p = -\rho$, stays constant despite the expansion. In this way, room for matter full of energy could have been created. The vacuum energy is converted into particles and radiation at some later stage and, in particular, the observed huge entropy is created. Potentially, this mechanism works for any inflationary scenario, since the product ρa^3 is guaranteed to grow whenever $\ddot{a} > 0$. However, the important question is whether a graceful exit out of the inflationary stage and successful reheating is possible. In practice, this issue has killed a number of inflationary models. Remarkably, the original model by A. Guth [86] had been ruled out precisely on these grounds [90].

Inflation has to continue for a sufficiently long time for the problems of horizon, curvature, and entropy to be solved. All these give roughly the same condition on the number of required 'e-foldings' of inflation [86] and we consider here a (simplified) derivation based on entropy. A precise condition can be found, for example, in Ref. [91]. Multiplying the current temperature in the Universe by its visible size we find $Ta\chi_0 \sim 10^{30}$, where χ_0 is the comoving size of the present horizon. The product Ta is conserved (up to the change in the number of relativistic degrees of freedom, which we neglect here) since the Universe expansion is adiabatic after the end of inflation, see Eq. (11). Let T_r denote the reheating temperature and $\mathrm{e}^N \equiv a_f/a_i$ the number of inflationary 'e-foldings', where a_f is the value of the scale factor at the end of inflation and a_i at its beginning, respectively. We also want at least the whole visible universe to be inflated out of a single causally connected patch, which gives $a_i\chi_0 \sim H_i^{-1}$, where H_i is the value of the Hubble parameter during inflation. All this gives the condition[1]

$$\frac{T_r}{H_i}\ \mathrm{e}^N \gtrsim 10^{30}\ . \tag{91}$$

In popular models of inflation the ratio T_r/H_i is within a couple orders of magnitude from unity, and we find $N \gtrsim 65$.

7.2 Models of inflation

Consider $T_{\mu\nu}$ for a scalar field φ

$$T_{\mu\nu} = \partial_\mu\varphi\,\partial_\nu\varphi - g_{\mu\nu}\,\mathcal{L} \tag{92}$$

with the Lagrangian :

$$\mathcal{L} = \partial_\mu\varphi\,\partial^\mu\varphi - V(\varphi)\ . \tag{93}$$

In a state when all derivatives of φ are zero, the stress-energy tensor of a scalar field simplifies to $T_\mu^\nu = V(\varphi)\,\delta_\mu^\nu$. This corresponds to a vacuum state. Indeed, comparing with Eq. (4), we find $V = \rho = -p$. There are two basic ways to arrange $\varphi \approx$ const and hence to imitate the vacuum-like state.

1. Consider the potential $V(\phi)$, which has a local minimum with a non-zero energy density separated from the true ground state by a potential barrier [86]. A universe which happened to be trapped in the meta-stable minimum will stay there for a while (since such a state can decay only via subbarrier tunnelling) and expansion of the universe will diminish all field gradients. Then the Universe enters a vacuum state. This model is ruled out since the inhomogeneities created during the phase transition that terminates the inflationary phase are too large [90]. However, the model is good for purposes of illustration. The frequently asked question is: How can it be that the energy density stays constant despite the expansion? In the model with local potential minimum the energy cannot decrease (classically) below the local minimum value, and therefore it has to stay constant despite the expansion.

[1] Strictly speaking in this relation T_r is not the real temperature in a state of thermal equilibrium, but $T_r \sim \rho_r^{1/4}$, where ρ_r is the energy density at the moment when the expansion becomes dominated by relativistic particles.

2. A. Linde was the first to realize that things work in the simplest possible setup [89]. Consider the potential

$$V(\phi) = \frac{1}{2} m_\phi^2 \phi^2 \, . \tag{94}$$

The equation of field motion in an expanding Universe is $\ddot{\phi} + 3H\dot{\phi} + m_\phi^2 \phi = 0$. If $H \gg m$, the 'friction' is too big and the field (almost) does not move. Therefore time derivatives in $T_{\mu\nu}$ can be neglected, and inflation starts (in a sufficiently homogeneous patch of the Universe). A Hubble parameter in this case is determined by the potential energy, $H \approx m\phi/M_{\rm Pl}$, and we see that inflation starts if the initial field value happens to satisfy $\phi > M_{\rm Pl}$. During inflationary stage the field slowly rolls down the potential hill. This motion is very important in the theory of structure creation; inflation ends when $\phi \sim M_{\rm Pl}$. At this time, field oscillations start around the potential minimum and later decay into radiation. In this way matter was likely created in our Universe.

7.3 Unified theory of creation

During inflation, and by its end, the Universe was in a vacuum-like state. We have to understand how this 'vacuum' was turned into the matter we observe around us, and how primordial fluctuations that gave rise to galaxies were created. Fortunately, these problems can be formalized in a nice and unified way. Basically, everything reduces to a problem of particle creation in a time-dependent classical background. On top of every 'vacuum' there are fluctuations of all quantum fields which are present in a given model. This bath of virtual quanta is indestructible, and even inflation cannot get rid of it. Being small, fluctuations of any field obey an oscillator equation

$$\ddot{u}_k \; + \; [k^2 + m_{\rm eff}^2(\eta)] \; u_k = 0 \, , \tag{95}$$

here u_k are amplitudes of fluctuating fields in Fourier space. Effective mass becomes time-dependent through the coupling to time-dependent background. Because $m_{\rm eff}$ is time-dependent, it is not possible to keep fluctuations in a vacuum. If one arranges to put oscillators with momentum k into the vacuum at one time, they will not be in vacuum at a later time because positive and negative frequency solutions mix, see below. Several remarks are in order.

- Equation (95) is valid for all particle species.
- The equation looks that simple in a conformal reference frame $ds^2 = a(\eta)^2 \; (d\eta^2 - dx^2)$. (Everywhere in this chapter a 'dot' means derivative with respect to η.)
- Of particular interest are ripples of space-time itself: curvature fluctuations (scalar fluctuations of the metric) and gravity waves (tensor fluctuations of the metric).
- The effective mass $m_{\rm eff}$ can be non-zero even for massless fields. Gravitational waves give the simplest example [92], with $m_{\rm eff}^2 = -\ddot{a}/a$. The effective mass for curvature fluctuations has a similar structure $m_{\rm eff}^2 = -\ddot{z}/z$, but with a being replaced by $z \equiv a\dot{\phi}/H$, see Refs. [93]– [96].
- For conformally coupled, but massive scalar, $m_{\rm eff} = m_0 \, a(\eta)$.

Note that creation was only possible because nature is not conformally invariant. Otherwise, $m_{\rm eff}$ is time-independent and a vacuum remains a vacuum forever. There are two important instances of a time-varying classical background in cosmology:

- Expansion of space-time, $a(\eta)$.
- Motion of the inflaton field, $\phi(\eta)$.

Both can be operational at any epoch of creation:

- During inflation (superhorizon size perturbations).
- While the inflaton oscillates (reheating).

During inflation superhorizon-sized perturbations of the metric are created, which give seeds for Large Scale Structure (LSS) formation and eventually lead to the formation of galaxies, and therefore of the Solar system and all the rest which we can see around us. During reheating, matter itself is created. Overall, there are four different situations (two sources times two epochs). If coupling to the inflaton is not essential, the corresponding process will be called 'pure gravitational creation' in what follows.

There are several primary observables which can be calculated out of u_k and further used for calculation of quantities of interest. Most useful are:

– The particle occupation numbers, n_k. Integration over d^3k gives the particle number density.

– The power spectrum of field fluctuations, $P_k \equiv u_k^* u_k$. Integration over d^3k gives the field variance.

Depending on physical situation, only one or the other may have sense. The particle number in a comoving volume is useful because it

– is adiabatic-invariant on sub-horizon scales (or when $m > H$);

– allows one to calculate abundances of various relics, e.g. dark matter.

But it has no meaning at super-horizon scales when $m < H$. The power spectrum and/or field variance is useful because it

– does not evolve on super-horizon scales if $m < H$;

– allows one to calculate density perturbations generated during inflation;

– is crucial for dynamics of phase transitions;

– helps to calculate back-reaction in a simple way (Hartree approximation).

But P_k evolves on sub-horizon scales and when $m > H$.

Let me start with a discussion of metric perturbations.

Gravitational creation of metric perturbations. As an important and simple example, let us consider quantum fluctuations of a real scalar field, which we denote as φ. It is appropriate to rescale the field values by the scale factor, $\varphi \equiv \phi/a(\eta)$. This brings the equations of motion for the field ϕ into a simple form of Eq. (95). As usual, we decompose ϕ over creation and annihilation operators $b_{\mathbf{k}}$ and $b^\dagger_{\mathbf{k}}$

$$\phi(\mathbf{x},\eta) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[u_k(\eta)\, b_{\mathbf{k}}\, \mathrm{e}^{i\mathbf{kx}} + u_k^*(\eta)\, b^\dagger_{\mathbf{k}}\, \mathrm{e}^{-i\mathbf{kx}}\right] . \tag{96}$$

The mode functions u_k satisfy Eq. (95). In what follows we will assume that φ is the inflaton field of the 'chaotic' inflationary model, Eq. (94). During inflation $H \gg m$ and $H \approx \mathrm{const}$. So, to start with, we can assume that φ is a massless field on the constant deSitter background. (The massive case can be treated similarly, but analytical expressions are somewhat more complicated and do not change the result in a significant way. Corrections due to change of H can also be taken into account, and we do that later for the purpose of comparison with observations.) With a constant Hubble parameter during inflation the solution of Friedmann equations in conformal time is

$$a(\eta) = -\frac{1}{H\eta} \tag{97}$$

and the equation for mode functions of a massless, conformally coupled to gravity ($\xi = 0$), scalar field takes the form

$$\ddot{u}_k + k^2 u_k - \frac{2}{\eta^2}\, u_k = 0 . \tag{98}$$

Solutions which start as vacuum fluctuations in the past ($\eta \to -\infty$) are given by

$$u_k = \frac{\mathrm{e}^{\pm ik\eta}}{\sqrt{2k}} \left(1 \pm \frac{i}{k\eta}\right) . \tag{99}$$

Indeed, at $\eta \to -\infty$ the second term in the parentheses can be neglected and we have the familiar mode functions of the Minkowski space–time. The wavelength of a given mode becomes equal to the horizon size (or 'crosses' the horizon) when $k\eta = 1$. Inflation proceeds with $\eta \to 0$, so the modes with progressively larger k cross the horizon. After horizon crossing, when $k\eta \ll 1$, the asymptotics of mode functions are

$$u_k = \pm \frac{i}{\sqrt{2}k^{3/2}\eta} , \quad \text{or} \quad \varphi_k = \frac{u_k}{a(\eta)} = \mp \frac{iH}{\sqrt{2}k^{3/2}} . \tag{100}$$

The field variance is given by

$$\langle 0|\phi^2(x)|0\rangle = \int \frac{d^3k}{(2\pi)^3} \, |\varphi_k|^2 , \tag{101}$$

and we find in the asymptotic (the careful reader will recognize that this is already a regularized expression with zero-point fluctuations being subtracted)

$$\langle \varphi^2 \rangle = \frac{H^2}{(2\pi)^2} \int \frac{dk}{k} . \tag{102}$$

Defining the power spectrum of the field fluctuations as a power per decade, $\langle \varphi^2 \rangle \equiv \int P_\varphi(k) \, d\ln k$, we find

$$P_\varphi(k) = \frac{H^2}{(2\pi)^2} . \tag{103}$$

Curvature perturbations. According to Eq. (3), the three-dimensional curvature of space sections of constant time is inversely proportional to the scale factor squared, ${}^{(3)}R \propto a^{-2}$. Therefore, the perturbation of spatial curvature is proportional to $\delta a/a$, and this ratio can be evaluated as

$$\zeta \equiv \frac{\delta a}{a} = H\delta t = H \, \frac{\delta\varphi}{\dot\varphi} . \tag{104}$$

This allows one to relate the power spectrum of curvature perturbations to the power spectrum of field fluctuations

$$P_\zeta(k) = \frac{H^2}{\dot\varphi^2} \; P_\varphi(k) , \tag{105}$$

and we find for the power spectrum of curvature perturbations

$$P_\zeta(k) = \frac{1}{4\pi^2} \, \frac{H^4}{\dot\varphi^2} . \tag{106}$$

This very important relation describes the inflationary creation of primordial perturbations, and can be confronted with observations. The usefulness of curvature perturbations for this procedure can be appreciated in the following way:

1. Consider the perturbed metric, Eq. (46). The product $a(1-\Phi)$ for the long-wavelength perturbations can be viewed as a perturbed scale factor, i.e. $\delta a/a = -\Phi$. Comparing this relation with Eq. (104) and Eq. (50), we find for the temperature fluctuations that are of superhorizon size at the surface of last scattering

$$\frac{\delta T}{T} = \frac{2}{3} \, \zeta_k . \tag{107}$$

2. On superhorizon scales the curvature perturbations do not evolve.[2] This fact allows one to relate directly the observed power spectrum of temperature fluctuations to the power spectrum of curvature fluctuations generated during inflation.

[2] I should warn you that this is quite a generic statement and holds in situations usually considered. Thus, it is forgotten sometimes that this is not a universally true statement. To avoid possible confusion when encountering specific complicated models, the reader should keep this fact in mind.

Tensor perturbations. Mode functions of gravity waves (after rescaling by $M_{\rm Pl}/\sqrt{32\pi}$) obey the same equation as mode functions of a massless minimally coupled scalar [92]. Using the result of Eq. (103) we immediately find [97]

$$P_T(k) = 2\,\frac{32\pi}{M_{\rm Pl}^2}\,P_\varphi(k) = \frac{16}{\pi}\,\frac{H^2}{M_{\rm Pl}^2}\,, \tag{108}$$

where the factor of 2 accounts for two graviton polarizations.

Slow-roll approximation. During inflation, the field φ rolls down the potential hill very slowly. A reasonable approximation to the dynamics is obtained by neglecting $\ddot{\varphi}$ in the field equation $\ddot{\varphi} + 3H\dot{\varphi} + V' = 0$. This procedure is called the slow-roll approximation

$$\dot{\varphi} \approx -\frac{V'}{3H}\,. \tag{109}$$

Field derivatives can also be neglected in the energy density of the inflaton field, $\rho \approx V$

$$H^2 = \frac{8\pi}{3M_{\rm Pl}^2}\,V\,. \tag{110}$$

This gives for curvature perturbations

$$\zeta_k \equiv P_\zeta(k)^{1/2} = \frac{H^2}{2\pi\,\dot{\varphi}} = \frac{4H}{M_{\rm Pl}^2}\,\frac{V}{V'}\,. \tag{111}$$

Normalizing to CMBR. As an example, let us consider the simplest model $V = \frac{1}{2}m^2\varphi^2$. We have

$$\frac{V}{V'} = \frac{\varphi}{2}, \qquad \text{and} \quad H = \sqrt{\frac{4\pi}{3}}\,\frac{m\varphi}{M_{\rm Pl}}\,. \tag{112}$$

This gives for the curvature fluctuations

$$\zeta_k = \sqrt{\frac{16\pi}{3}}\,\frac{m\,\varphi^2}{M_{\rm Pl}^3}\,. \tag{113}$$

Using the relation between curvature and temperature fluctuations, Eq. (107), and normalizing $\delta T/T$ to the measured value at largest l, which is $\delta T/T \sim 10^{-5}$ (see Fig. 5), we find the restriction on the value of the inflaton mass in this model:

$$m \approx \frac{\delta T}{T}\,\frac{M_{\rm Pl}}{30} \approx 10^{13}\ \ {\rm GeV}\,. \tag{114}$$

Here I have used the fact that in this model the observable scales cross the horizon when $\varphi \approx M_{\rm Pl}$.

Slow-roll parameters. The number of e-foldings ($a = {\rm e}^{Ht} \equiv {\rm e}^N$) of inflationary expansion from the time when $\varphi = \varphi_i$ to the end can be found as

$$N(\varphi_i) = \int_{t_i}^{t_f} H(t)dt = \int \frac{H}{\dot{\varphi}}\,d\varphi = \frac{8\pi}{M_{\rm Pl}^2}\int_{\varphi_e}^{\varphi_i}\frac{V}{V'}\,d\varphi\,. \tag{115}$$

In particular, in the model Eq. (94) we find that the largest observable scale crossed the horizon ($N \sim 65$) when $\varphi_i \approx 3.5M_{\rm Pl}$. All cosmological scales which fit within the observable universe encompass a small $\Delta\phi$ interval within $M_{\rm Pl} < \varphi < \varphi_i$. And the inflaton potential should be sufficiently flat over this range of $\Delta\phi$ for inflation to proceed. This means that observables essentially depend on the first few derivatives

of V (in addition the the potential $V(\phi_0)$ itself). From the first two derivatives one can construct the following dimensionless combinations

$$\epsilon \equiv \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2 , \tag{116}$$

$$\eta \equiv \frac{M_{\rm Pl}^2}{8\pi}\frac{V''}{V} , \tag{117}$$

which are often called the slow-roll parameters.

The power spectra of curvature, Eq. (105), and of tensor perturbations, Eq. (108), in slow-roll parameters can be rewritten as

$$P_\zeta(k) = \frac{1}{\pi\epsilon}\frac{H^2}{M_{\rm Pl}^2} , \qquad P_T(k) = \frac{16}{\pi}\frac{H^2}{M_{\rm Pl}^2} . \tag{118}$$

Comparing these two expressions we find

$$\frac{P_T(k)}{P_\zeta(k)} = 16\epsilon . \tag{119}$$

Primordial spectrum. In general, the spectra can be approximated as power law functions in k:

$$P_\zeta(k) = P_\zeta(k_0)\left(\frac{k}{k_0}\right)^{n_S-1} , \tag{120}$$

$$P_T(k) = P_T(k_0)\left(\frac{k}{k_0}\right)^{n_T} . \tag{121}$$

To the first approximation, H in Eq. (118) is constant. Therefore, in this approximation, power spectra do not depend on k and $n_S = 1$, $n_T = 0$. This case is called the Harrison–Zel'dovich spectrum [98, 99] of primordial perturbations. However, in reality, H is changing, and in Eq (118) for every k one should take the value of H at the moment when the relevant mode crosses the horizon. In slow-roll parameters one then finds (see, for example, Ref. [100] for a nice overview)

$$n_S = 1 + 2\eta - 6\epsilon, \qquad n_T = -2\epsilon . \tag{122}$$

We can re-write Eq. (123) as a relation between the slope of tensor perturbations and the ratio of power in tensor to curvature modes

$$\frac{P_T(k)}{P_\zeta(k)} = -8n_T . \tag{123}$$

This is called the *consistency relation* that (simple) inflationary models should obey.

Different models of inflation have different values of slow-roll parameters η and ϵ, and therefore can be represented in the (η,ϵ) parameter plane. Using the relations in Eq. (122) we see that this plane can be mapped into (n_S, n_T), or using also Eq. (123) into the (n_S, r) parameter plane, where r is the ratio of power in tensor to scalar (curvature) perturbations. In this way, different inflationary models can be linked to observations and constraints can be obtained.

The most recent constraints in the (n_s, r) plane, utilizing WMAP and SDSS data, are presented in the left panel of Fig. 20. The shaded dark red region is ruled out by WMAP alone. The shaded light red region is ruled out when adding SDSS information. The two dotted lines delimit the three classes of inflationary models known as the small-field, large-field and hybrid models. Some single-field models of inflation make highly specific predictions in this plane, as indicated. From top to bottom, the figure shows the predictions for $V(\phi) \propto \phi^6$ (line segment; ruled out by CMB alone), $V(\phi) \propto \phi^4$ (star; on verge of exclusion) and $V(\phi) \propto \phi^2$ (line segment; the inflation model Eq. (94); still allowed).

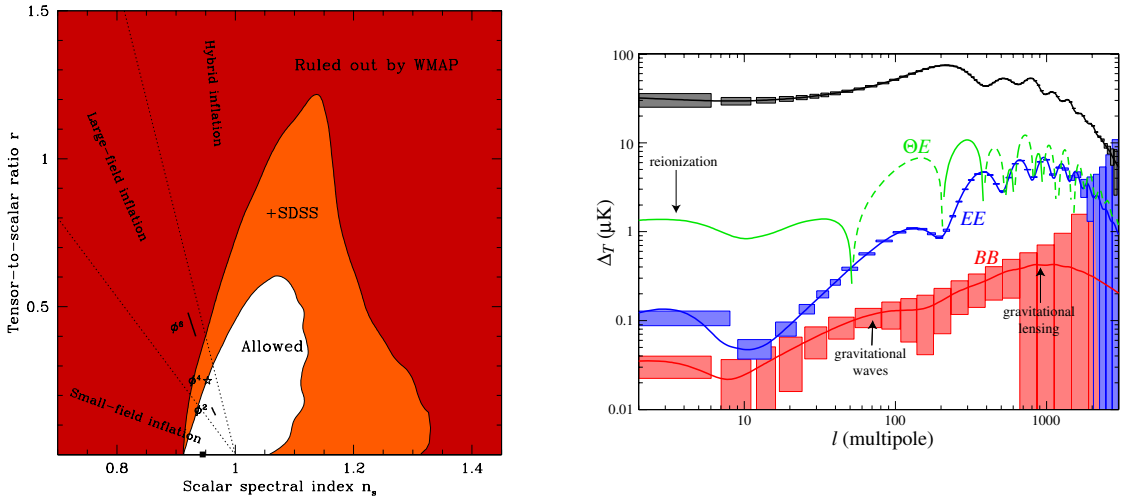


Fig. 20: Left panel: 95% constraints on inflationary models in the (n_s, r) plane. From Ref. [10]. Right panel: Forecast for the planned ESA Planck mission.

Testing inflation. All predictions of inflationary cosmology which could be tested so far, have been confirmed. In particular, the Universe is spatially flat (within experimental errors), $\Omega = 1.02 \pm 0.02$, see Table 1. The primordial perturbations are of superhorizon size and adiabatic. The spectral index is close to the Harrison–Zeldovich case, see Fig. 20, left panel. A crucial test of inflationary models would be the detection of gravity waves and verification of the consistency relation. These signatures of typical inflationary models are within reach of future CMBR experiments, see Fig. 20, right panel.

8 ULTRA-HIGH-ENERGY COSMIC RAYS

In the early years, cosmic-ray studies were ahead of accelerator research, starting from the discovery of positrons, through muons, to that of pions and strange particles. Today we are facing the situation that the puzzling saga of cosmic rays of the highest energies may again unfold in the discovery of new physics, now beyond the Standard Model; or it may bring to life an 'extreme' astrophysics.

Immediately after the discovery of the relict Cosmic Microwave Background Radiation (CMBR), Greisen, Zatsepin and Kuzmin [101,102] realized that the highest-energy protons should catastrophically lose energy in photo-production of pions on this universal background. This process limits the distance to the farthest sources of observed rays to be roughly 100 Mpc and should lead to the cut-off in the energy spectrum. However, the number of events with energies beyond the expected cut-off as measured by different installations is growing with time [103]– [109], while no nearby sources where identified. The findings of Greisen, Zatsepin and Kuzmin (GZK) are based on solid fundamental physics which involve precisely measured cross-sections in a GeV energy range (in the centre-of-mass reference frame). Therefore, if the data are correct—and it is believed they are basically correct[3]—one should either invoke new physics, or accept unconventional and uncomfortable very 'extreme' astrophysics. This is the reason for the excitement and growing interest in ultra-high-energy cosmic-ray research; for recent reviews see Refs. [110]– [113].

Methods of detection. At energies below 10^{14} eV, the flux of cosmic rays is sufficiently high that direct measurements using high-altitude balloons or satellite experiments are possible. Above 10^{15} eV, the flux is only one particle per square metre per year, which excludes direct observations on the orbit.

[3]While a disagreement in measured fluxes has recently emerged, there is no reason to doubt the reality of super-GZK events.

At 10^{20} eV the number is down to one particle per square kilometre per century. Here the problem for direct measurements would be not only a vanishingly small flux, but the enormously high energy itself. (Remember that calorimeters at modern colliders weigh hundreds of thousands of tonnes.) Fortunately, the major part of our UHECR detectors is already built for us by Nature and is put, rotating, into orbit: the Earth's atmosphere makes a perfect calorimeter. The atmosphere is just thick enough so that showers of secondary particles produced by incoming cosmic rays of the highest energies, in collisions with nuclei of air, reach their maximum intensity just above the Earth's surface. Particles in a shower propagate with the velocity of light, forming a thin disk perpendicular to the direction of the incident particle. At 10^{19} eV the footprint of the shower on the ground is several kilometres across.

The shower can be registered either by placing an array of particle detectors on the Earth's surface, or by measuring the Cherenkov light produced by particles in the atmosphere, or by tracking the fluorescence light emitted when shower particles excite nitrogen molecules in the air. Particle detectors in a ground array can be spaced hundreds of metres apart and are operational around the clock. Fluorescence light telescopes see the cosmic-ray track just like a fly's eye would see the meteorite, but only moving with the speed of light. These detectors are operational only on clear moonless nights, but are able to measure the longitudinal shower profile and its maximum depth directly.

With either technique, the energy and incident direction of primary particle can be measured shower by shower. Chemical composition also can be inferred, but only in a statistical sense, after averaging over many showers.

1. *Arrival direction.* The timing of a signal in different detectors is used to determine the direction of a shower (ground-array technique). Direction is measured with an accuracy of about 2°. The measurement is straightforward and does not involve any uncertainties. Inferred information is reliable. Fluorescence light telescopes observe the whole shower track, and in stereo mode the precision of angle determination is 0.5°.

2. *Energy.* The energy estimate, on the other hand, is not that straightforward. In fluorescent light detectors, the energy of the primary particle is derived from the observed light intensity, therefore incorrect modelling and/or insufficiently frequent monitoring of atmosphere transparency can be a source of errors. For the ground-array detectors, the energy estimate relies on a Monte Carlo model of shower development and is related to the shower density profile. Nevertheless, the currently favoured model, QGSJET [114], describes data well from TeV up to highest energies and it is believed that the overall error (statistical plus systematic) in energy determination does not exceed 30%.

The best would be to employ both the ground array and fluorescent light techniques simultaneously. This should reduce systematic errors, and this is the design of the forthcoming Pierre Auger project.

3. *Chemical composition.* Chemical composition can be inferred from the details of shower development. For example, showers initiated by heavy nuclei start earlier in the atmosphere and have fewer fluctuations compared to proton showers. Fluorescence detectors observe shower development directly. Using ground-array detectors, the shower depth can be extracted by measuring, for example, the ratio of electrons to muons. At lowest energies, the chemical composition of cosmic rays reflects primary and secondary element abundances; for a recent review see Ref. [115]; at highest energies, $E > 4 \times 10^{19}$ eV, the conclusion is that less than 50% of primary cosmic rays can be photonic at 95% confidence level [116].

8.1 Propagation of ultra-high-energy cosmic rays

In this subsection, we consider the influence of different cosmological backgrounds on the propagation of highest-energy cosmic rays.

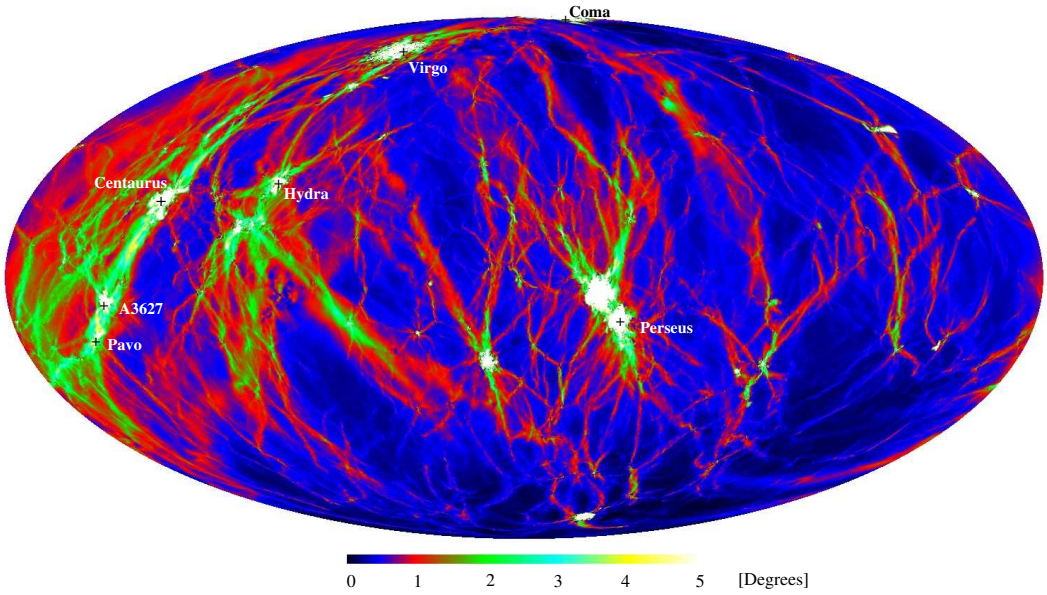


Fig. 21: Full sky map of deflection angles for UHECRs with energy 4×10^{19} eV after travelling 100 Mpc in an extra-galactic magnetic field. The coordinate system is galactic, with the galactic anti-centres in the middle of the map. Positions of identified clusters are marked using the locations of the corresponding halos in the simulation. The map is obtained in a magneto-hydrodynamical simulation of cosmic structure formation that correctly reproduces the positions and masses of known galaxy clusters in the Local Universe. From Ref. [117].

Magnetic fields. Magnetic fields play an important role in the processes of cosmic-ray acceleration and propagation, their trajectories being bent by the action of the Lorentz force

$$\frac{d\vec{v}}{dt} = \frac{Ze}{E}\ [\vec{v}\times\vec{B}]\ . \tag{124}$$

For a qualitative discussion, it is often sufficient to compare a gyro-radius of the trajectory of a relativistic particle

$$R_g = \frac{E}{ZeB} \tag{125}$$

to other relevant length scales of the problem. For example, a magnetic 'trap' can not confine a cosmic ray if the gyro-radius exceeds the trap size. The deflection angle $\Delta\theta$, after traversing the distance L in a homogeneous magnetic field, is proportional to $L/R_{\rm L}$. In a chaotic magnetic field, the deflection angle will grow as $\sqrt{L}$. Let us estimate a typical deflection angle, L/R_g, of a charged UHE particle after traversing Galactic or extra-galactic magnetic fields.

1. In the Galactic magnetic field, for particles coming across the Galactic disc, we have

$$\frac{\Delta\theta}{Z} \approx 2.5^\circ\ \frac{10^{20}\ {\rm eV}}{E}\ \frac{B}{3\ \mu{\rm G}}\ \frac{L}{1.5\ {\rm kpc}}\ , \tag{126}$$

where $3\ \mu$G is the magnitude of the regular magnetic field and 1.5 kpc is the width of the disc. We see that protons with $E > 10^{18}$ eV escape our Galactic disc easily. Protons of smaller energy are trapped and can escape the Galaxy only by diffusion and 'leaking' from the boundary. Cosmic rays with $E > 10^{18}$ eV should be extra-galactic, if protons. Even if cosmic rays would be all iron nuclei, at $E > 2\cdot10^{19}$ eV cosmic rays should be extra-galactic, otherwise strong anisotropy in the direction of the Galactic disc would have been observed.

2. Extra-galactic magnetic fields have not yet been measured, except for the central regions of galaxy clusters [118]. However, there is an upper bound on their strength from the (absence of) Faraday rotation of polarized extra-galactic sources [119, 120]. This translates to the upper bound on deflections in extra-galactic magnetic fields

$$\frac{\Delta\theta}{Z} < 2.5^\circ \ \frac{10^{20}\ \mathrm{eV}}{E} \ \frac{B}{10^{-9}\ \mathrm{G}} \ \frac{(L\lambda)^{1/2}}{10\ \mathrm{Mpc}} \ , \tag{127}$$

where λ is the coherence length of the extra-galactic magnetic fields; λ is believed to be smaller than 1 Mpc. However, extragalactic fields are strongly inhomogeneous, with amplitude changing by orders of magnitude from clusters to filaments, and from filaments to voids. Deflections in some directions, which do not cross clusters and strong filaments, may be small, otherwise deflections can be very large. This situation cannot realistically be described by a mean field.

Only recently have attempts been made to simulate UHECR propagation in a realistically structured universe [117, 121]. Results from Ref. [117] are shown in Fig. 21. Additional motivation for this simulation was to obtain, in constraint simulations of the Local Structure, a realistic map of expected deflections, which would reflect the positions of known clusters. Such a map can be used in the analysis of cosmic-ray arrival directions. Resulting deflections do not exceed the resolution of UHECR experiments over most of the sky. About an order of magnitude stronger deflections were obtained in Ref. [121]. There are two possible reasons for disagreement. First, simulations of Ref. [121] were unconstrained and therefore do not reflect our concrete local neighbourhood. Second, variable resolution of Ref. [117] was better in cluster regions, which is a possible reason for the larger obtained dynamical range between fields in clusters and filaments. Since in both simulations the magnetic fields are normalized to typical values in the core of rich clusters, their values in the filaments will be very different, with larger fields outside clusters in Ref. [121]. Work aimed at resolving these differences is in progress. For now, I adopt the results of Ref. [117] and conclude that arrival directions of UHECR should point back to the sources. Charged particle astronomy of UHECR is, in principle, possible.

Interactions with cosmic-radiation backgrounds. Ultra-high energy cosmic rays cannot propagate elastically in cosmic backgrounds. They have enough energy to produce secondary particles even in collisions with CMBR (important for proton primaries) or radio photons (important for UHECR photons) or infrared radiation (important for propagation of nuclei). Most important is the reaction of pion photoproduction for protons (or neutrons) propagating in relic cosmic microwave background left over from the Hot Big Bang. For the threshold energy of this reaction we find, in the laboratory frame,

$$E_{\mathrm{th}}(p+\gamma \to N+\pi) = \frac{(m_p+m_\pi)^2 - m_p^2}{2E_\gamma(1-\cos\theta)} \ . \tag{128}$$

Note that in the derivation of this relations, standard Lorentz kinematic and standard dispersion relations between particle energy and momentum, $E^2 = k^2 + m^2$, are assumed. If any of these are violated, the threshold condition in a laboratory frame may look different. For the black-body distribution of CMBR photons with temperature $T = 2.7$ K we find

$$E_{\mathrm{th}} \approx 5 \times 10^{19}\ \mathrm{eV} \ . \tag{129}$$

This reaction has a large cross-section, being the largest at the Δ resonance. At half-width of the resonance

$$\sigma \sim 300\ \mu\mathrm{b} \approx 3 \times 10^{-28}\ \mathrm{cm}^2 \ . \tag{130}$$

Density of CMB radiation is $n \sim T^3 \sim 400\ \mathrm{cm}^{-3}$. This corresponds to the mean free path:

$$L_\sigma = (\sigma n)^{-1} \approx 8 \times 10^{24}\ \mathrm{cm} \approx 2.7\ \mathrm{Mpc} \ . \tag{131}$$

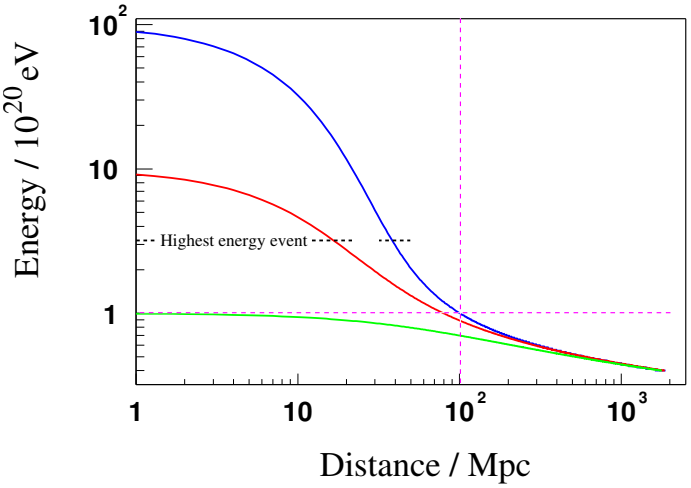


Fig. 22: Energy of protons as a function of the distance propagated in CMBR for three initial values of energy at the source, 10^{22}, 10^{21} and 10^{20} eV respectively.

In each collision $\approx 20\%$ of energy is lost (which is the mass of a pion). Successive collisions rob protons of energy, which decreases exponentially. The distance over which energy decreases by one e-fold is called the attenuation length. At the threshold, Eq. (129), the attenuation length is large, $L_{\rm A} \approx 10^3$ Mpc (being determined by other processes, see below.) With increasing energy, it rapidly decreases and at energies above the Δ resonance for typical CMBR photons, $E \gtrsim 5 \cdot 10^{20}$ eV, the attenuation length is $L_{\rm A} \approx 10$ Mpc. It follows that the energy of protons drops below 10^{20} eV after it travels the distance of order 100 Mpc almost independently of initial energy, see Fig. (22). We conclude that

[A1] *Protons detected with* $E > 10^{20}$ eV *should have an origin within* $R < R_{\rm GZK} \equiv 100$ Mpc.

We will call the corresponding volume a GZK sphere (or GZK distance).

The reaction $p + \gamma \to p + e^+e^-$ is sub-dominant. While it has a smaller threshold (by a factor of $2m_e/m_\pi \sim 10^{-2}$), it also has a smaller cross-section. But, it becomes important at sub-GZK energies. The attenuation length for this reaction is 10^3 Mpc—a noticeable and important effect.

UHE photons lose energy in the $\gamma + \gamma \to e^+e^-$ reaction. The threshold for the reaction with CMBR photons is smaller by a factor of $2m_e^2/m_\pi m_p \sim 10^{-5}$ compared to the GZK cutoff energy. The cross-section decreases fast with energy, $\sigma = \sigma_{\rm T} m_e/s^2$, where $\sqrt{s}$ is the CM energy and $\sigma_{\rm T} \approx 10^{-22}{\rm cm}^2$ is the Thomson cross-section. Therefore, the attenuation length has a minimum at the pair-production threshold. For CMBR photons, this occurs at $E \approx 2 \cdot 10^{14}$ eV and $L_{\rm A} \approx 10$ kpc. The attenuation length increases with energy, reaching a GZK distance roughly at $E \approx 10^{20}$ eV. Photons with even larger energies are able to penetrate even larger distances—and this is important for many models—but in this energy range, the main contribution comes from poorly known radio-background, which brings some uncertainty in the attenuation length of the highest energy photons.

Heavy nuclei lose energy in photo-dissociation. Here, the main contribution comes from the infrared background which is also poorly known. But again, at $E \approx 10^{20}$ eV the attenuation length is comparable to the GZK distance [122].

The cutoff. It is easy to understand why a sharp cutoff in the spectrum of protons should appear. This happens because the attenuation length decreases rapidly with increasing energy. Assume a power law injection spectrum for UHECR, $J_{\rm in}(E) \propto E^{-\alpha}$, and let $n(r)$ be the density of sources. Fluxes from individual sources decrease as r^{-2}, which is compensated by volume integration, $r^2 dr$. Therefore, the

total flux registered at energy E should grow in proportion to the upper limit of volume integration

$$J(E) \propto \int_0^{R(E)} n(r)dr \ \propto \ R(E) \ , \tag{132}$$

if the distribution of sources, $n(r)$, does not depend on r. Here, $R(E)$ corresponds to the attenuation length, i.e. the distance from which cosmic rays with energy E can reach us. The attenuation length of protons with $E < 5 \times 10^{19}$ eV equals 10^3 Mpc, while the attenuation length at $E > 5 \cdot 10^{20}$ eV is only 10 Mpc. We conclude that

[A2] *The drop in flux by 2 orders of magnitude at GZK energy is expected if the distribution of sources is homogeneous.*

A word of caution is needed here. Transition in $R(E)$ from the sub-GZK to the super-GZK regime is not instantaneous. Therefore, a particular value of the drop depends upon the shape of the injection spectrum, i.e. on the value of α, see, for example, Refs. [123]– [125].

8.2 Generation of UHECR

The origin of cosmic rays and/or their acceleration mechanisms has been a subject of debate for several decades. Particles can be accelerated either by astrophysical shock waves, or by electric fields. In either case, one can estimate the maximum energy; with optimistic assumptions, the final estimate is the same for both mechanisms. In practice, the maximum energy is expected to be much lower.

1. *Shock acceleration.* Particles are accelerated stochastically while bouncing between shocks. Acceleration can continue only if particles remain confined within an accelerating region, in other words until gyro-radius, Eq. (125), is smaller than the size of the region. This gives

$$E_{\max} = ZeBL \ . \tag{133}$$

2. *Acceleration by an electric field.* The latter can be created by a rapidly rotating magnetized neutron star or black hole. If motion is relativistic, the generated electric field is of the same order as the magnetic field, and the difference in electric potentials is $\sim (B \times L)$. This, again, reproduces Eq. (133) for the maximum energy.

Known astrophysical sources with $(B \times L)$ big enough to give $E_{\max} \sim 10^{20}$ eV are neutron stars, active galactic nuclei (AGNs) and colliding galaxies.

The central engine of an AGN is believed to be a super-massive black hole powered by matter accretion. AGNs have two jets (one of the jets may be invisible because of the Doppler effect) of relativistic particles streaming in opposite directions. Interaction with the intergalactic medium terminates this motion and the radio lobes and hot spots are formed at the ends of jets, see Fig. 23. The acceleration of UHECR primaries may occur either near the black hole horizon (direct acceleration), or in hot spots of jets and radio lobes (shock acceleration). A host of different AGNs is now classified in one unified scheme, for a review see Ref. [126]. Depending upon the angle between the jet axis and the line of sight, we observe different types of AGN. A typical radio galaxy, showing two strong opposite jets, is observed at angles approximately perpendicular to the jet axis. An AGN is classified as a quasar if the angle is smaller than 30°. If we look along the jet axis (angle $< 10^\circ$), i.e. directly into the barrel of the gun, we observe a blazar.

It should be noted that not all radio galaxies are the same. There are Fanaroff–Riley (FR) type I and type II galaxies (radio-loud AGNs), and Seyfert galaxies (radio-quiet AGNs). Both types of FR galaxies may be the sites of UHECR acceleration, but the hot spots in FR type II galaxies are considered to be the most promising [128]. It is believed that when observed along the jet axis, FR type II galaxies make a parent population of Highly Polarized Quasars (HPQ—subclass of blazars), while FR type I galaxies make a parent population of BL Lacertae objects (BL Lacs—another subclass of blazars).

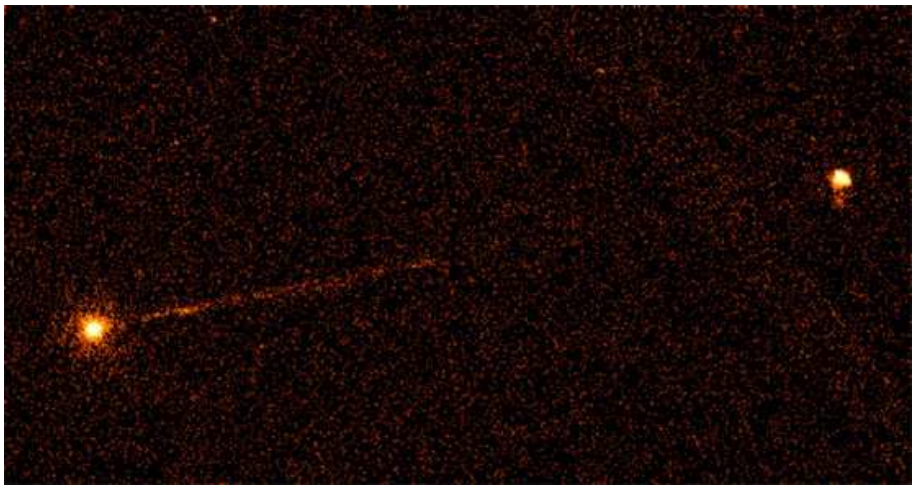


Fig. 23: Chandra telescope X-ray image of the nucleus, jet, and hot spot of Pictor A. From Ref. [127].

As an example, the X-ray image of the powerful FR type II radio galaxy Pictor A taken by the Chandra X-ray Observatory is shown in Fig. 23. Radio observations of jets have a long history. Recently, Chandra started to obtain high resolution X-ray maps of AGNs which, surprisingly, revealed very long collimated X-ray jets. For example, the distance from the nucleus to the hot spot in Pictor A is at least 240 kpc. It is hard to explain such long jets as pure leptonic, and it is possible that the population of relativistic electrons responsible for the X-ray emission is a result of photo-pion production by UHE protons [127].

Now, for any acceleration mechanism and independent of the actual acceleration site (i.e. be it either the AGN's black hole or any of the hot spots), the momentum of highest energy particles is expected to point in the direction of the jet [129]. In other words, if AGNs are sources of UHECR, arrival directions of high-energy cosmic rays may point back to a (subclass) of a blazar family. Such correlations were indeed observed [129]– [132] with BL Lacertae objects.

8.3 UHECR spectrum

The largest statistic of UHECR events has been accumulated for over 12 years of operation by the AGASA air shower array of surface particle detectors. The spectrum measured by AGASA is shown in Fig. 24, left panel. The dotted curve represents the theoretical expectation for a homogeneous distribution of sources and proton primaries. This theoretical curve exhibits the GZK cut-off at $E \approx 10^{20}$ eV. Remarkably, AGASA has detected 11 events with higher energy and the data show no hint for cutoff.

It is hard to argue against the reality of these findings. AGASA exposure is under control, and the only issue is the energy determination. AGASA events have an accuracy of $\pm 25\%$ in event-reconstruction resolution and 18% in systematic errors around 10^{20} eV [108]. Added in quadrature this gives r.m.s. error of energy determination to be $\pm 30\%$. More importantly, the probability of an upward fluctuation to 1.5 times the true energy is 2.8%. There are too many super-GZK events, and with this resolution a spectrum with GZK cutoff cannot be transformed into an excess of post-GZK events assuming spillover [112].

Recently, the HiRes group has reported results obtained with a telescope which measures atmospheric fluorescence light. The energy spectrum is shown in Fig. 24 by triangles, right panel. The spectrum is consistent with the GZK cutoff, and there are two events detected with $E > 10^{20}$ eV. Systematic error in energy measurement was estimated to be 21%, systematic error in the aperture is not yet clear. HiRes employs a relatively new technique, with the following issues usually cited for improve-

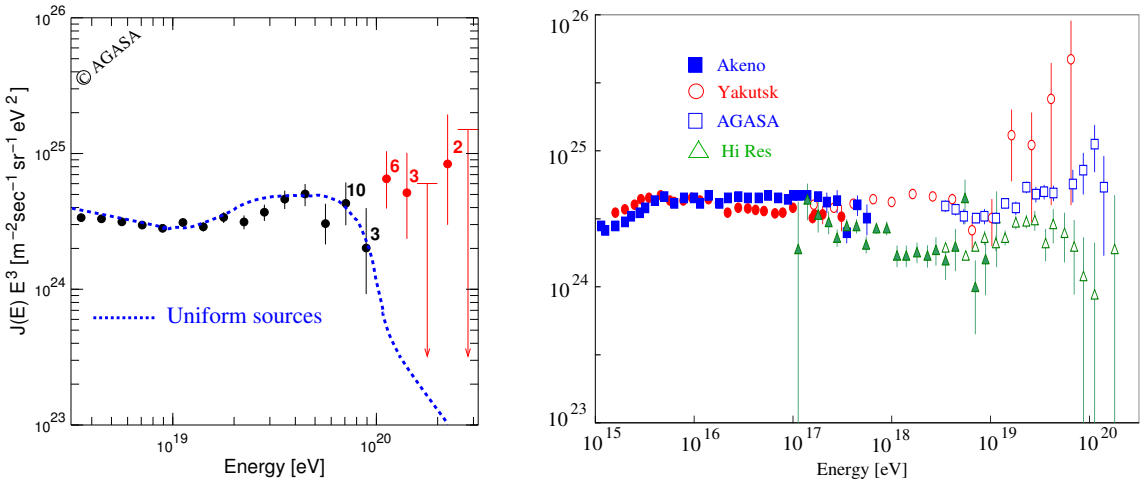


Fig. 24: **Left panel:** The energy spectrum of cosmic rays with a zenith angle up to 45° as measured by AGASA [108]. Numbers near data points reflect the number of events in the respective energy bins. **Right panel:** A compilation of data from different experiments: Akeno (filled squares), AGASA (open squares), HiRes-I and HiRes-II (open and filled triangles), two Yakutsk sub-arrays of Cherenkov detectors (open and filled circles). From Ref. [133].

ment: atmospheric attenuation corrections should be based on nightly measurements and not averages, better energy calibration and aperture calculation are called for, see, for example, Ref. [112].

The Yakutsk group uses a hybrid detection method, combining a ground array of particle detectors with telescopes that measure Cherenkov light produced by a shower in the atmosphere. A recently reported [133] spectrum, derived from air Cherenkov light measurements, is shown in the right panel of Fig. 24 by circles. At the low-energy end it agrees well with the Akeno spectrum, and at the high-energy end it is consistent, within errors, with the AGASA spectrum. Yakutsk and AGASA disagree significantly with HiRes at $E \sim 10^{18}$ eV where statistical errors are negligible, which points to some systematics. AGASA and HiRes can be reconciled at $E < 10^{20}$ by, for example, a -15% and $+15\%$ respective shift of energy [134]. The discrepancy between two experiments at $E > 10^{20}$ after the shift is only 2σ. However, even after these shifts, there are still nine events with $E > 10^{20}$ in the combined data set.

8.4 The puzzle

These measurements are regarded as a threefold puzzle, because contrary to assertions [A1] and [A2]

[P1] no candidate sources are found within the GZK distance in the directions of $E > 10^{20}$ eV events;

[P2] the AGASA spectrum does not exhibit the GZK cutoff.

And finally we have the third puzzling question:

[P3] Which physical processes are capable of producing events with these enormous energies?

Conjectures. With the assumption that all three pieces of the puzzle, [P1]–[P3], are correct, the situation becomes desperate. There were no solutions suggested which would not invoke new physics beyond the Standard Model, or very speculative astrophysics. In addition, all models require fine tuning, and many do not really solve all three problems. It is not possible to consider here all the suggested solutions. Ignoring for now problem [P2], the situation with [P1] and [P3] does not become easier, but we can now restrict ourselves to the discussion of astrophysical solutions only.

Ignoring in addition [P3], the simplest suggestion is to assume very large extragalactic magnetic fields, which would randomize UHECR trajectories. However, the results of Ref. [117] do not support such a conjecture. As we have mentioned already, a consensus regarding EGMF fields has not yet been reached and in Ref. [121], much stronger extragalactic magnetic fields were advocated. Nevertheless, even in this case, the conclusion was that the condition of global isotropy of UHECR arrival directions requires the 'local' value of the magnetic field to be rather weak, $B \lesssim 0.1\ \mu\mathrm{G}$, which, in turn, leads to a large number of UHECR sources in the GZK volume, $n \gtrsim 100$; for similar limits on the number of sources see also Refs. [135]– [138]. These weak EGMFs of Refs. [117, 121] rule out the possibility of a single powerful radio-galaxy, which happened to be nearby [139], or a gamma-ray burst scenario [140], as potential sources of UHECR.

Another suggested astrophysical scenario was a 'dead quasars' model [141]. This model assumes that quasars, powerful in the past, retain the possibility to accelerate to the highest energies even after the accretion of matter is exhausted and a quasar ceases to be visible electromagnetically. However, the process of acceleration to the highest energies in compact sources is inevitably accompanied by a strong TeV emission [142]. Recent results obtained by several TeV telescopes, in particular, non-observation of strong TeV sources, rule out the 'dead quasar' model [143]. In addition, in Ref. [144] it was found that known quasar remnants are typically distributed too anisotropically to explain the isotropic ultra-high-energy cosmic-ray flux except in the unrealistic case where extragalactic magnetic fields of $0.1\ \mu$G extend over many Mpc.

A possibility that ultra-high-energy events are due to iron nuclei accelerated from young, strongly magnetized neutron stars in relativistic MHD winds has also been suggested [145]. However, with realistic parameters of Galactic magnetic field, even iron nuclei do not propagate diffusively within the Galaxy, which disfavours this model [146].

Any observational clue? Many quite different models were suggested for the resolution of the GZK puzzle. The majority of suggested models, which we have no space to consider here, employ a new physics of one sort or another. (The reader may consult UHECR reviews cited at the beginning of this section, but I believe that a review which would cover all the suggested possibilities does not exist.) Instead, let us consider the question of whether or not there is already a clue in the data as to which model may be correct. Hints, and, in principle, critical signatures are given by:

- *Spectral shape.* We do not yet have enough data at the highest energies to constrain models. Spectra below 10^{20} eV point to the AGN model of UHECR origin, with protons being primaries [147, 148].
- *Chemical composition.* Again, there is not enough data at the highest energies. An analysis of Haverah Park data at lower energies shows that above 10^{19} eV, less than 30% of the primary cosmic rays can be photons or iron nuclei at the 95% confidence level [149]. In other words, at least 70% should be protons.
- *Large-scale anisotropy.* Gives strong signatures. Not observed, which is a hint by itself. We have considered some implications already, and may add that the non-observation of anisotropy towards the Galactic centre has a potential of ruling out the model of UHECR origin based on decays of super-heavy dark matter [150]– [153].
- *Small-scale clustering.* This is an observed [154], reliable feature. (Errors in angle determination are definitely small.) It is already statistically significant. Therefore, below I shall concentrate on this signature.

Small-scale clustering. It was observed by different installations that arrival directions of UHECR are too close to each other and this happens too often [155]– [157]. In particular, the AGASA collaboration has observed six doublets and one triplet of cosmic rays with $E > 4 \times 10^{19}$ within 2.5°. The chance probability of observing just a triplet under an isotropic distribution is only 0.9% [154]. The statistical

significance of these clusters in the AGASA data set was considered by several authors. In Ref. [157], an analysis based on the calculation of an angular autocorrelation function was employed, and the probability $P = 3 \times 10^{-4}$ of chance clustering was obtained. This includes the penalty for the choice of the energy cut, while the angular bin was chosen to be fixed at 2.5°, which is a value previously accepted by AGASA, being consistent with the angular resolution. In Ref. [158], this analysis was repeated and confirmed. In addition, two more conservative estimates were made. In the first, the penalty factor for the adjustment of the angular bin size was added. This returns $P = 3 \times 10^{-3}$. This is a valid procedure, but it misses prior information about the angular resolution of the installation. In the second estimate, the bin size was kept fixed, but the whole data set was divided in halves. The 'original data set' [155] was used to justify the bin size of 2.5°, while clusters in it were removed for the subsequent evaluation of statistical significance. This procedure returned $P = 8 \times 10^{-2}$. Again, this too is a valid approach, and can be safely used with future large data sets. However, I'd like to stress that it is *not* an evaluation of the statistical significance of six doublets and one triplet. It is no wonder that a smaller data set has reduced statistical significance. Finally, in Ref. [159] it was found that the AGASA data set manifests a $P \sim 10^{-3}$ chance probability of clustering above background using independent statistics of $\langle \cos\theta \rangle_{[0^\circ, 10^\circ]}$. I find this value, $P \sim 10^{-3}$, to be a fair estimate of the current significance of clustering in the AGASA data.

Note the following: if clusters are real and due to sources, the number of events in 'physical' clusters should have a Poisson distribution. Therefore, with the current low statistics, it is expected that roughly half of installations should observe significant clustering, while another half should not see it [157]. There is no clustering in the current HiRes data [159, 160]. However, with the current statistics there is no contradiction yet [159, 161].

The study of small-scale clustering is very important. If clusters are real and not a statistical fluctuation, then UHECR should point back to sources and UHECR astronomy is possible. Real sources should be behind the clusters and the correlation studies make sense. Pursuing this strategy, one should be restricted to astrophysical sources with physical conditions potentially suitable for particle acceleration to the highest energies. Active galactic nuclei (AGN) constitute a particularly attractive class of potential sources. As we have already discussed, if AGNs are sources, those which have jets directed along the line of sight, or blazars, should correlate with observed UHECR events. It is intriguing that statistically significant correlations of UHECR with BL Lacertae objects were found [129].

9 CONCLUSIONS

Cosmology and astrophysics give us firm evidence that the Standard Model of particle physics is limited. The Standard Model fails to explain baryogenesis, does not contain non-baryonic dark matter, and has no room for massive neutrinos. We now know that dark energy also exists, but we do not know why it exists. There seem to be too many coincidences between numerical values of cosmological parameters that describe the matter and energy budget. Contributions of baryonic matter, non-baryonic matter, and dark energy are almost equal at the present epoch, while they have seemingly unrelated origin and could differ by many orders of magnitude. Cosmology just became a precision science and is already full of surprises; we can expect even more exciting discoveries in the near future.

ACKNOWLEDGEMENTS

I would like to thank the conference organizers for the friendly and warm atmosphere.

REFERENCES

[1] G. Gabadadze, Beyond the Standard Model, *in* 2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, 2003.

[2] S. Petcov, Massive neutrinos and neutrino oscillations, *in* 2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, 2003.

[3] A. Chilingarian, Galactic and solar cosmic rays, *in* 2003 European School of High-Energy Physics, Tsakhkadzor, Armenia, 2003.

[4] V. Rubakov, Cosmology and astrophysics,*in* Proc. 2001 European School of High-Energy Physics, Beatenberg, Switzerland, 2001.

[5] M. Shaposhnikov, Cosmology and astrophysics, *in* Proc. 2000 European School of High-Energy Physics, Caramulo, Portugal, 2000.

[6] J. Garcia-Bellido, Astrophysics and cosmology, *in* Proc. 1999 European School of High-Energy Physics, Casta-Papernicka, Slovak Republic, 1999, hep-ph/0004188.

[7] K. A. Olive, G. Steigman and T. P. Walker, Phys. Rep. **333**, 389 (2000), astro-ph/9905320.

[8] V. A. Rubakov and M. E. Shaposhnikov, Usp. Fiz. Nauk **166**, 493 (1996), hep-ph/9603208.

[9] C. L. Bennett *et al.*, Astrophys. J. Suppl. **148**, 1 (2003), astro-ph/0302207.

[10] SDSS, M. Tegmark *et al.*, (2003), astro-ph/0310723.

[11] A. G. Riess *et al.*, (2004), astro-ph/0402512.

[12] D. J. Schwarz, Annalen Phys. **12**, 220 (2003), astro-ph/0303574.

[13] B. M. S. Hansen *et al.*, Astrophys. J. **574**, L155 (2002).

[14] W. L. Freedman *et al.*, Astrophys. J. **553**, 47 (2001), astro-ph/0012376.

[15] A. A. Penzias and R. W. Wilson, Astrophys. J. **142**, 419 (1965).

[16] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, Astrophys. J. **142**, 414 (1965).

[17] T. Shmaonov, Prib. Tekh. Ehksp. **1**, 83 (1957).

[18] A. McKellar, Proc. Astron. Soc. Pac. **52**, 187 (1940).

[19] A. G. Doroshkevich and I. D. Novikov, Sov. Phys. Dokl. **9**, 111 (1964).

[20] D. J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996), astro-ph/9605054.

[21] M. Kamionkowski and L. Knox, Phys. Rev. **D67**, 063001 (2003), astro-ph/0210165.

[22] C. L. Bennett *et al.*, Astrophys. J. **464**, L1 (1996), astro-ph/9601067.

[23] G. F. Smoot *et al.*, Astrophys. J. **396**, L1 (1992).

[24] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).

[25] J. Silk, Astrophys. J. **151**, 459 (1968).

[26] P. J. E. Peebles and J. T. Yu, Astrophys. J. **162**, 815 (1970).

[27] R. A. Syunyaev and Y. B. Zel'Dovich, Astrophys. Space Sci. **7**, 3 (1970).

[28] A. G. Doroshkevich, Y. B. Zel'Dovich, and R. A. Syunyaev, Sov. Astron. **22**, 523 (1978).

[29] W. Hu and S. Dodelson, Annu. Rev. Astron. Astrophys. **40**, 171 (2002), astro-ph/0110414.

[30] W. Hu and M. J. White, Phys. Rev. Lett. **77**, 1687 (1996), astro-ph/9602020.

[31] F. R. Bouchet, P. Peter, A. Riazuelo, and M. Sakellariadou, Phys. Rev. **D65**, 021301 (2002), astro-ph/0005022.

[32] D. Huterer and M. S. Turner, Phys. Rev. **D64**, 123527 (2001), astro-ph/0012510.

[33] Supernova Search Team, A. G. Riess *et al.*, Astron. J. **116**, 1009 (1998), astro-ph/9805201.

[34] D. N. Spergel *et al.*, Astrophys. J. Suppl. **148**, 175 (2003), astro-ph/0302209.

[35] G. Soucail, J. P. Kneib, and G. Golse, (2004), astro-ph/0402658.

[36] E. Corbelli and P. Salucci, (1999), astro-ph/9909252.

[37] K. G. Begeman, A. H. Broeils, and R. H. Sanders, Mon. Not. R. Astron. Soc. **249**, 523 (1991).

[38] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490**, 493 (1997).

[39] S. Ghigna *et al.*, Astrophys. J. **544**, 616 (2000), astro-ph/9910166.

[40] P. Sikivie, I. I. Tkachev, and Y. Wang, Phys. Rev. **D56**, 1863 (1997), astro-ph/9609022.

[41] G. Steigman and I. Tkachev, (1998), astro-ph/9803008.

[42] W. H. Kinney and P. Sikivie, Phys. Rev. **D61**, 087305 (2000), astro-ph/9906049.
[43] C. Charmousis, V. Onemli, Z. Qiu, and P. Sikivie, Phys. Rev. **D67**, 103502 (2003), astro-ph/0301399.
[44] D. P. Bennett *et al.*, Astrophys. J. **579**, 639 (2002), astro-ph/0109467.
[45] EROS, C. Afonso *et al.*, Astron. Astrophys. **400**, 951 (2003), astro-ph/0212176.
[46] P. Popowski *et al.*, (2003), astro-ph/0304464.
[47] J. Yoo, J. Chaname, and A. Gould, Astrophys. J. **601**, 311 (2004), astro-ph/0307437.
[48] F. Zwicky, Helv. Phys. Acta **6**, 110 (1933).
[49] A. D. Lewis, D. A. Buote, and J. T. Stocke, Astrophys. J. **586**, 135 (2003), astro-ph/0209205.
[50] R. Carlberg *et al.*, Astrophys. J. **462**, 32 (1996), astro-ph/9509034.
[51] M. Loewenstein and R. Mushotzky, (2002), astro-ph/0208090.
[52] J. P. Kneib, R. S. Ellis, I. Smail, W. J. Couch, and R. M. Sharples, Astrophys. J. **471**, 643 (1996), astro-ph/9511015.
[53] M. Bartelmann and P. Schneider, Phys. Rep. **340**, 291 (2001), astro-ph/9912508.
[54] G. Squires *et al.*, Astrophys. J. **461**, 572 (1996), astro-ph/9507008.
[55] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[56] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[57] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[58] G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
[59] R. Bradley *et al.*, Rev. Mod. Phys. **75**, 777 (2003).
[60] A. D. Dolgov, Phys. Rep. **370**, 333 (2002), hep-ph/0202122.
[61] G. G. Raffelt, Annu. Rev. Nucl. Part. Sci. **49**, 163 (1999), hep-ph/9903472.
[62] S. I. Blinnikov and M. Y. Khlopov, Sov. J. Nucl. Phys. **36**, 472 (1982).
[63] Z. Berezhiani, D. Comelli, and F. L. Villante, Phys. Lett. **B503**, 362 (2001), hep-ph/0008105.
[64] R. Foot and R. R. Volkas, Phys. Rev. **D68**, 021304 (2003), hep-ph/0304261.
[65] K. A. Olive, (2003), astro-ph/0301505.
[66] J. R. Ellis, (2003), astro-ph/0305038.
[67] P. Gondolo, (2004), astro-ph/0403064.
[68] V. Berezinsky, M. Kachelriess, and A. Vilenkin, Phys. Rev. Lett. **79**, 4302 (1997), astro-ph/9708217.
[69] V. A. Kuzmin and V. A. Rubakov, Phys. Atom. Nucl. **61**, 1028 (1998), astro-ph/9709187.
[70] D. J. H. Chung, E. W. Kolb, and A. Riotto, Phys. Rev. **D59**, 023501 (1999), hep-ph/9802238.
[71] V. Kuzmin and I. Tkachev, JETP Lett. **68**, 271 (1998), hep-ph/9802304.
[72] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996), hep-ph/9506380.
[73] DAMA, R. Bernabei *et al.*, Phys. Lett. **B480**, 23 (2000).
[74] CDMS, D. S. Akerib *et al.*, (2004), astro-ph/0405033.
[75] S. S. Gershtein and Y. B. Zeldovich, JETP Lett. **4**, 120 (1966).
[76] G. Bhattacharyya, H. Paes, L.-g. Song, and T. J. Weiler, Phys. Lett. **B564**, 175 (2003), hep-ph/0302191.
[77] S. Tremaine and J. E. Gunn, Phys. Rev. Lett. **42**, 407 (1979).
[78] K. Griest and M. Kamionkowski, Phys. Rev. Lett. **64**, 615 (1990).
[79] C. J. Hogan and M. J. Rees, Phys. Lett. **B205**, 228 (1988).
[80] E. W. Kolb and I. I. Tkachev, Phys. Rev. Lett. **71**, 3051 (1993), hep-ph/9303313.
[81] E. W. Kolb and I. I. Tkachev, Astrophys. J. **460**, L25 (1996), astro-ph/9510043.
[82] L. F. Abbott and P. Sikivie, Phys. Lett. **B120**, 133 (1983).

[83] M. Dine and W. Fischler, Phys. Lett. **B120**, 137 (1983).
[84] V. Kuzmin and I. Tkachev, Phys. Rev. **D59**, 123006 (1999), hep-ph/9809547.
[85] A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).
[86] A. H. Guth, Phys. Rev. **D23**, 347 (1981).
[87] A. D. Linde, Phys. Lett. **B108**, 389 (1982).
[88] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[89] A. D. Linde, Phys. Lett. **B129**, 177 (1983).
[90] A. H. Guth and E. J. Weinberg, Phys. Rev. **D23**, 876 (1981).
[91] A. R. Liddle and D. H. Lyth, Phys. Rep. **231**, 1 (1993), astro-ph/9303019.
[92] L. P. Grishchuk, Sov. Phys. JETP **40**, 409 (1975).
[93] V. N. Lukash, Sov. Phys. JETP **52**, 807 (1980).
[94] M. Sasaki, Prog. Theor. Phys. **76**, 1036 (1986).
[95] V. F. Mukhanov, Sov. Phys. JETP **67**, 1297 (1988).
[96] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rept. **215**, 203 (1992).
[97] V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. **B115**, 189 (1982).
[98] E. R. Harrison, Phys. Rev. **D1**, 2726 (1970).
[99] Y. B. Zeldovich, Mon. Not. R. Astron. Soc. **160**, 1 (1972).
[100] J. E. Lidsey *et al.*, Rev. Mod. Phys. **69**, 373 (1997), astro-ph/9508078.
[101] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
[102] G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4**, 78 (1966).
[103] J. Linsley, Phys. Rev. Lett. **10**, 146 (1963).
[104] M. M. Winn, J. Ulrichs, L. S. Peak, C. B. A. Mccusker, and L. Horton, J. Phys. **G12**, 653 (1986).
[105] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, J. Phys. **G17**, 733 (1991).
[106] A. V. Glushkov *et al.*, Bull. Russ. Acad. Sci. Phys. **55**, No. 495 (1991).
[107] D. J. Bird *et al.*, Astrophys. J. **441**, 144 (1995).
[108] M. Takeda *et al.*, Astropart. Phys. **19**, 447 (2003), astro-ph/0209422.
[109] High Resolution Fly's Eye, T. Abu-Zayyad *et al.*, (2002), astro-ph/0208301.
[110] M. Nagano and A. A. Watson, Rev. Mod. Phys. **72**, 689 (2000).
[111] L. Anchordoqui, T. Paul, S. Reucroft, and J. Swain, Int. J. Mod. Phys. **A18**, 2229 (2003), hep-ph/0206072.
[112] J. W. Cronin, (2004), astro-ph/0402487.
[113] D. F. Torres and L. A. Anchordoqui, (2004), astro-ph/0402371.
[114] N. N. Kalmykov and S. S. Ostapchenko, Phys. Atom. Nucl. **56**, 346 (1993).
[115] K.-H. Kampert, Heavy Ion Phys. **14**, 203 (2001), astro-ph/0101331.
[116] M. Ave, J. A. Hinton, R. A. Vazquez, A. A. Watson, and E. Zas, Phys. Rev. **D65**, 063007 (2002), astro-ph/0110613.
[117] K. Dolag, D. Grasso, V. Springel, and I. Tkachev, (2003), astro-ph/0310902.
[118] C. L. Carilli and G. B. Taylor, Annu. Rev. Astron. Astrophys. **40**, 319 (2002), astro-ph/0110655.
[119] P. P. Kronberg, Rep. Prog. Phys. **57**, 325 (1994).
[120] P. Blasi, S. Burles, and A. V. Olinto, Astrophys. J. **514**, L79 (1999), astro-ph/9812487.
[121] G. Sigl, F. Miniati, and T. A. Ensslin, (2004), astro-ph/0401084.
[122] F. W. Stecker and M. H. Salamon, Astrophys. J. **512**, 521 (1999), astro-ph/9808110.
[123] M. Blanton, P. Blasi, and A. V. Olinto, Astropart. Phys. **15**, 275 (2001), astro-ph/0009466.
[124] O. E. Kalashev, V. A. Kuzmin, D. V. Semikoz, and I. I. Tkachev, (2001), astro-ph/0107130.
[125] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, (2001), hep-ph/0107306.

[126] C. M. Urry and P. Padovani, Publ. Astron. Soc. Pac. **107**, 803 (1995), astro-ph/9506063.
[127] A. S. Wilson, A. J. Young, and P. L. Shopbell, Astrophys. J. **547**, 740 (2001), astro-ph/0008467.
[128] J. P. Rachen and P. L. Biermann, Astron. Astrophys. **272**, 161 (1993), astro-ph/9301010.
[129] P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74**, 445 (2001), astro-ph/0102476.
[130] P. G. Tinyakov and I. I. Tkachev, Astropart. Phys. **18**, 165 (2002), astro-ph/0111305.
[131] D. S. Gorbunov, P. G. Tinyakov, I. I. Tkachev, and S. V. Troitsky, Astrophys. J. **577**, L93 (2002), astro-ph/0204360.
[132] P. Tinyakov and I. Tkachev, (2003), astro-ph/0305363.
[133] A. A. Ivanov, S. P. Knurenko, and I. Y. Sleptsov, Nucl. Phys. Proc. Suppl. **122**, 226 (2003), astro-ph/0305053.
[134] D. De Marco, P. Blasi, and A. V. Olinto, Astropart. Phys. **20**, 53 (2003), astro-ph/0301497.
[135] S. L. Dubovsky, P. G. Tinyakov, and I. I. Tkachev, Phys. Rev. Lett. **85**, 1154 (2000), astro-ph/0001317.
[136] Z. Fodor and S. D. Katz, Phys. Rev. **D63**, 023002 (2001), hep-ph/0007158.
[137] P. Blasi and D. De Marco, Astropart. Phys. **20**, 559 (2004), astro-ph/0307067.
[138] D. Harari, S. Mollerach, and E. Roulet, (2004), astro-ph/0404304.
[139] G. R. Farrar and T. Piran, (2000), astro-ph/0010370.
[140] E. Waxman, Phys. Rev. Lett. **75**, 386 (1995), astro-ph/9505082.
[141] E. Boldt and P. Ghosh, Mon. Not. R. Astron. Soc. **307**, 491 (1999), astro-ph/9902342.
[142] A. Levinson, Phys. Rev. Lett. **85**, 912 (2000).
[143] A. Neronov, P. Tinyakov, and I. Tkachev, (2004), astro-ph/0402132.
[144] C. Isola, G. Sigl, and G. Bertone, (2003), astro-ph/0312374.
[145] P. Blasi, R. I. Epstein, and A. V. Olinto, Astrophys. J. **533**, L123 (2000), astro-ph/9912240.
[146] S. O'Neill, A. V. Olinto, and P. Blasi, (2001), astro-ph/0108401.
[147] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, (2002), hep-ph/0204357.
[148] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, (2002), astro-ph/0210095.
[149] M. Ave, J. A. Hinton, R. A. Vazquez, A. A. Watson, and E. Zas, Phys. Rev. Lett. **85**, 2244 (2000), astro-ph/0007386.
[150] S. L. Dubovsky and P. G. Tinyakov, JETP Lett. **68**, 107 (1998), hep-ph/9802382.
[151] V. Berezinsky and A. A. Mikhailov, Phys. Lett. **B449**, 237 (1999), astro-ph/9810277.
[152] H. B. Kim and P. Tinyakov, (2003), astro-ph/0306413.
[153] M. Kachelriess and D. V. Semikoz, Phys. Lett. **B577**, 1 (2003), astro-ph/0306282.
[154] AGASA, M. Takeda *et al.*, Astrophys. J. **522**, 225 (1999), astro-ph/9902239.
[155] AGASA, N. Hayashida *et al.*, Phys. Rev. Lett. **77**, 1000 (1996).
[156] Y. Uchihori *et al.*, Astropart. Phys. **13**, 151 (2000), astro-ph/9908193.
[157] P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74**, 1 (2001), astro-ph/0102101.
[158] C. B. Finley and S. Westerhoff, (2003), astro-ph/0309159.
[159] The High Resolution Fly's Eye (HIRES), R. U. Abbasi *et al.*, (2004), astro-ph/0404366.
[160] The High Resolution Fly's Eye (HIRES), R. U. Abbasi *et al.*, (2004), astro-ph/0404137.
[161] H. Yoshiguchi, S. Nagataki, and K. Sato, (2004), astro-ph/0404411.

ORGANIZING COMMITTEE

George Pogosyan (JINR and ICAS, Yerevan State University), School Director
Tatyana Donskova (JINR), Organizing Secretary
Nick Ellis (CERN)
Robert Fleischer (CERN)
Aram Kotzinyan (Yerevan State University, JINR, and CERN)
Egil Lillestøl (CERN and University of Bergen)
Levon Mardoyan (ICAS, Yerevan State University)
Danielle Métral (CERN), Organizing Secretary
Alexander Olchevski (JINR and CERN)
Alexei Sissakian (JINR)

LECTURERS

Ian Aitchison (University of Oxford)
Ashot Chilingarian (YerPhI, Armenia)
Robert Fleischer (CERN)
Gregory Gabadadze (New York University)
Alexander Khodjamirian (University of Karlsruhe)
Iosef Manjavidze (JINR)
Serguey Petcov (SISSA, Trieste)
Igor Tkachev (CERN)

DISCUSSION LEADERS

Laura Covi (DESY)
Dmitry Fursaev (JINR)
Alexei Pivovarov (INR, Moscow)
Tilman Plehn (CERN)
Michael Pluemacher (CERN)
Oleg Teryaev (JINR)

OTHER ATTENDEES

Vladimir Kadyshevsky, Director-General, JINR
Alexei Sissakian, Deputy Director-General, JINR
Roger Cashmore, Representing CERN Director-General
Armen Yeranyan, Assistant, Yerevan State University
Vardges Avagyan, Medical Doctor

STUDENTS

Mathieu Agelou
Giorgi Arabidze
Evgueni Baldine
Konstantin Batkov
Stéphanie Beauceron
Olaf Behrendt
Ralf Patrick Bernhard
Meta Binder
Ilija Bizjak
Ingo Bloch
Carmela Carpentieri
Antoine Cazes
Alexis Cothenet
Peter Cwetanski
Rita De Masi
Biagio Di Micco
Mauro Dinardo
Eric Dumonteil
Luigi Salvatore Esposito
Henning Flaecher
Bruce Gallop
Olivier Gaumer
Chiara Genta
Nerses Gevorgyan
Lukasz Goscilo
Sébastien Greder
Amber Jenkins
Josef Juran
Gia Khoriauli
Andrei Kiriakov
Alexander Klimenko
Aleksei Korablev
Venelin Kozhuharov
Marcin Kucharczyk
Anne-Catherine Le Bihan
Marie Legendre
Anne-Marie Magnan
Katharina Mair
Ewa Markiewicz
Adam Matyja
Niels Meyer
Alessandro Montanari
Carsten Noeding
Emily Nurse
Matthew Palmer
Charles Pattison
Roman Pekarik
Ciro Pistillo
Michael Poettgens
Marco Poli Lener
Robindra Prabhu
Melisa Rossi
Luca Scotto Lavina
Michal Secansky
Radomir Smida
Dmytro Soroka
Ivan Sotsky
Krzysztof Syryczynski
Silvano Tosi
Piotr Traczyk
Roberto Versaci
Sébastien Viret
Tuan Vu Anh
Ronald Weber
Viatcheslav Zholobov
Stéphanie Zimmermann

LISTENERS

Andranik Hakobyan
Vigen Kojoyan
Marianna Kuznetsova
Bakur Parsamyan

LIST OF POSTERS

Authors	**Poster title**
Evgueni Baldine	New precision measurement of the J/Ψ and Ψ' meson masses (produced with keDR Collaboration)
Konstantin Batkov	Simulating energy response of the ATIC calorimeter (produced with the ATIC Collaboration)
Simon Baumgartner	2HDM at HERA
Stéphanie Beauceron	$\mathrm{Wb\bar{b}}$ production at D0
Olaf Behrendt	Measurement of F_2 at low Q^2 with the H1 detector
Meta Binder	Search for SUSY in the $\mathrm{p\bar{p}} \rightarrow \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow \mu\mu\mu(\mathrm{e}) + X$ channel
Ilija Bizjak	Measurement of $\lvert V_{ub}\rvert$ on a fully reconstructed sample
Ingo Bloch	The ZEUS experiment at HERA
Carmelia Carpentieri	Semi-automatic assembly of back-to-back silicon detector modules for the ATLAS experiment
Antoine Cazes	Neutrino oscillations
Peter Cwetanski	Acceptance tests of the ATLAS TRT end-cap wheels (development of a gas monitor and straw tube simulations)
Rita De Masi	Silicon detectors in the COMPASS experiment
Biagio Di Micco	KLOE — A window in 1 GeV physics
Mauro Dinardo	BTeV — B physics at Tevatron, Fermilab
Eric Dumonteil	Some hardware and offline developments of the dimuon arm on the ALICE experiment at CERN
Luigi Salvatore Esposito	OPERA: RID and $\nu_\pi \rightarrow \nu_\mathrm{e}$ oscillation
Henning Flaecher	Hadronic mass moments from semileptonic B meson decays at BaBar
Bruce Gallop	System for ATLAS SCT module macro-assembly testing
Olivier Gaumer	ATLAS and $\mathrm{Z}' \rightarrow \mathrm{l}^-\mathrm{l}^+$ physics
Chiara Genta	*see Ingo Bloch*
Nils Gollub	Search for a heavy charged Higgs boson at ATLAS in the $\mathrm{gg} \rightarrow \mathrm{H}^{\pm}\mathrm{tb} \rightarrow \mathrm{tbtb}$ channel
Sebastien Greder *et al.*	Evidence of mini SUZY and maxi-excitation
Amber Jenkins	$\mathrm{Zb\bar{b}}$ Physics at D0
Gia Khoriauli	Tile calorimeter calibration for ATLAS
Andrei Kiriakov	SVD experiments search $\mathrm{c\bar{c}}$ for 70 GeV/*c*
Alexander Klimenko *et al.*	Timing studies of long scintillator counters for TOF detectors
Aleksei Korablev	Test of electromagnetic T-invariance in $\mathrm{K} \rightarrow \pi^0\mathrm{e}\nu\gamma$ decays

Authors	**Poster title**
Anne-Catherine Le Bihan	Tau identification in hadronic final states at D0 Run II
Marie Legendre	Study of time-dependent CP violation with partial reconstruction of $\mathrm{B}^0 \to \mathrm{D}^{*\pm}\pi^{\pm}$
Anne-Marie Magnan	D^0 searches for supersymmetric particles in a R-parity violation context with data taken by the Tevatron, Fermilab
Katharina Mair	ATLAS muon test-beam activities
Niels Meyer	Higgs width measurements with TESLA
Alessandro Montanari	*see Ingo Bloch*
Carsten Noeding	Supersymmetry search in channels with multileptons at D0
Emily Nurse	A measurement of σBr for $\mathrm{Z} \to \mu^+\mu^-$ at D0
Matthew Palmer	ATLAS SCT data acquisition software calibration control and data analysis — heavy vector boson decays of massive graviton resonances
Ciro Pistillo *et al.*	OPERA: an appearance experiment to search for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations in the CNGS beam
Michael Poettgens *et al.*	Test of CMS tracker silicon detector modules with the ARC readout system
Marco Poli Lener	Performance of a triple-GEN detector for high-rate particle triggering
Melisa Rossi	The CDF Run II experiment
Luca Scotto Lavina *et al.*	Status and goal of the OPERA experiment for the search of neutrino oscillations
Dmytro Soroka	Participation of Kharkov CMS Group in the CERN/CMS Monte Carlo event production
Ivan Sotsky	The study of four-lepton production at different polarizations of initial photons
Silvano Tosi	The BaBar experiment at SLAC
Roberto Versaci	*see Biagio Di Micco*
Sebastien Viret	B-physics in ATLAS: the $\mathrm{B_d} \to \mathrm{K}^{*0}\gamma$ example
Tuan Vu Anh	*See Carsten Noeding*
Ronald Weber	CMS pixel detector for LHC
Stephanie Zimmermann	*see Katharina Mair*