© Copyright CERN, Genève, 2003

Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document ; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.

ISSN 0007-8328
ISBN 92-9083-211-8
2001 CERN–CLAF SCHOOL OF HIGH-ENERGY PHYSICS

Itacuruçá, Brazil
6–19 May 2001

PROCEEDINGS
Eds: N. Ellis, B. Maréchal
Abstract

The CERN–CLAF School of High-Energy Physics is intended to give young physicists an introduction to the theoretical aspects of recent advances in elementary-particle physics. These proceedings contain lectures on field theory for the Standard Model, quantum chromodynamics, physics beyond the Standard Model, and neutrino physics, as well as reports on heavy-ion physics, instrumentation, and trigger and data-acquisition systems.
Preface

The first in the new series of Latin American Schools of High-Energy Physics organized jointly by CERN and CLAF (Centro Latino-Americano de Física), Rio de Janeiro, took place from 6 to 19 May 2001 at Itaúruca, Brazil with Bernard Marechal acting as local School director. Seventy-one students attended the inaugural School, 56 of them coming from eight Latin-American countries (17 from Mexico, 16 from Brazil, 11 from Argentina and 12 from other countries), 13 from Europe and two from the US. The 12 lecturers came from Europe, Latin America and the US. The lectures, which were in English, were complemented by daily discussion sessions led by seven physicists from Latin America. The students presented their work with enthusiasm in a poster session. The hotel and its location in Itaúruca were very well suited to the school, and the participants appreciated the beautiful surroundings.

The School was generously supported by Mexican funding agencies, Spain, Portugal, France and Italy, in addition to CERN and CLAF. Local support was provided by Brazilian agencies (CNPq, CAPES, FAPERJ, FUJB). Our sincere thanks go to all the sponsors making it possible to organize the School and contributing 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.

The major responsibility for running the School was in the hands of professor Bernard Marechal assisted by a local committee. Our special thanks go to Bernard, whose friendly attitude and omnipresence at the School contributed in major ways to the success of the School.

We are particularly indebted to Cintia Rodrigues da Cunha, the owner of the hotel, and to her enthusiastic hotel staff who in addition to normal hotel duties took part in more or less everything including musical assistance to the samba musician. Our thanks also go to the travel agency, Bon Voyage, for their valuable assistance in organizing most of the air travel and the local transport practically without any assistance from the school organizers.

We are very grateful to Claire Earnshaw for her efforts in the lengthy preparations for the school. Our greatest thanks go to Danielle Metral who stepped in at the last minute and took care of the day-to-day running of the School. Her efficient work, friendly attitude and continuous care of the students and their needs were highly appreciated.

The students will certainly remember several interesting excursions, which included a rainy visit to Rio and a fantastic excursion and picnic by the hotel’s own boat.

However, the success of the 2001 School was to a large extent due to the students themselves. Their poster session was very well prepared and highly appreciated, and throughout the School they participated actively during the lectures, in the discussion sessions, and in the different activities and excursions.

Egil Lillestøl
on behalf of the Organizing Committee
# List of people in photograph

<table>
<thead>
<tr>
<th>Family name</th>
<th>First name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alday</td>
<td>Luis Fernando</td>
<td>52</td>
</tr>
<tr>
<td>Alves</td>
<td>Gilvan</td>
<td>65</td>
</tr>
<tr>
<td>Arístizabal Sierra</td>
<td>Diego</td>
<td>81</td>
</tr>
<tr>
<td>Arteaga-Velazquez</td>
<td>Juan Carlos</td>
<td>41</td>
</tr>
<tr>
<td>Astorga</td>
<td>Patricio</td>
<td>27</td>
</tr>
<tr>
<td>Barbosa</td>
<td>Henrique de Melo Jorge</td>
<td>21</td>
</tr>
<tr>
<td>Barkovich</td>
<td>Mateo Alejandro</td>
<td>77</td>
</tr>
<tr>
<td>Barr</td>
<td>Alan</td>
<td>15</td>
</tr>
<tr>
<td>Betemps</td>
<td>Marcos</td>
<td>64</td>
</tr>
<tr>
<td>Boogert</td>
<td>Stewart</td>
<td>80</td>
</tr>
<tr>
<td>Brockill</td>
<td>Patrick John</td>
<td>3</td>
</tr>
<tr>
<td>Brodet</td>
<td>Eyal</td>
<td>61</td>
</tr>
<tr>
<td>Candia</td>
<td>Julian</td>
<td>40</td>
</tr>
<tr>
<td>Careno</td>
<td>Alexandra Stephanie</td>
<td>18</td>
</tr>
<tr>
<td>Catalaní</td>
<td>Fernando</td>
<td>9</td>
</tr>
<tr>
<td>Chavez</td>
<td>Helder</td>
<td>16</td>
</tr>
<tr>
<td>Cohen</td>
<td>Andy</td>
<td>58</td>
</tr>
<tr>
<td>Contreras</td>
<td>Guillermo</td>
<td>84</td>
</tr>
<tr>
<td>Cordero Cid</td>
<td>Luz Adriana</td>
<td>44</td>
</tr>
<tr>
<td>Cuattle Flores</td>
<td>Eleazar</td>
<td>17</td>
</tr>
<tr>
<td>da Silva</td>
<td>Roberto</td>
<td>55</td>
</tr>
<tr>
<td>Daleo</td>
<td>Alejandro</td>
<td>63</td>
</tr>
<tr>
<td>De Campos</td>
<td>Ana Helena</td>
<td>35</td>
</tr>
<tr>
<td>D’Hondt</td>
<td>Jorgen</td>
<td>50</td>
</tr>
<tr>
<td>Diaz</td>
<td>Marco Aurelio</td>
<td>43</td>
</tr>
<tr>
<td>Diaz Sanchez</td>
<td>Rodolfo Alexander</td>
<td>31</td>
</tr>
<tr>
<td>Dova</td>
<td>Maria Teresa</td>
<td>25</td>
</tr>
<tr>
<td>Ellis</td>
<td>Nick</td>
<td>51</td>
</tr>
<tr>
<td>Gallegos</td>
<td>Armando</td>
<td>30</td>
</tr>
<tr>
<td>Garcia</td>
<td>Carlos A.</td>
<td>54</td>
</tr>
<tr>
<td>Garcia Reyes</td>
<td>Enrique</td>
<td>82</td>
</tr>
<tr>
<td>Gonzalez</td>
<td>Javier</td>
<td>5</td>
</tr>
<tr>
<td>Grijpink</td>
<td>Sjors</td>
<td>70</td>
</tr>
<tr>
<td>Gwenlan</td>
<td>Claire</td>
<td>46</td>
</tr>
<tr>
<td>Hernandez-Mora</td>
<td>Hugo R.</td>
<td>11</td>
</tr>
<tr>
<td>Hernandez-Sanchez</td>
<td>Jaime</td>
<td>42</td>
</tr>
<tr>
<td>Herrera Herrera</td>
<td>Simon Fernando</td>
<td>1</td>
</tr>
<tr>
<td>Hesketh</td>
<td>Gavin</td>
<td>45</td>
</tr>
<tr>
<td>Liljestad</td>
<td>Egil</td>
<td>60</td>
</tr>
<tr>
<td>Lima</td>
<td>Ana</td>
<td>33</td>
</tr>
<tr>
<td>Lipari</td>
<td>Paolo</td>
<td>14</td>
</tr>
<tr>
<td>Luna</td>
<td>Rene</td>
<td>12</td>
</tr>
<tr>
<td>Machado</td>
<td>Ana Amelia</td>
<td>24</td>
</tr>
<tr>
<td>Mackedanz</td>
<td>Luiz Fernando</td>
<td>79</td>
</tr>
<tr>
<td>Marechal</td>
<td>Bernard</td>
<td>26</td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Number</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Mariazi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinez-Castro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medina-Gaytan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesquita</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moreno</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moura (de Moura Junior)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moura Santos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moyse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navarro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paganini</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pavololotzky</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelinson</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penny</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polycarpo (Ribeiro Polycarpo Macedo)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ponciano</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Razeira</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riquer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodriguez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodriguez Rodriguez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roman Lopez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roulet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanchez</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santos (Nunes Casimiro dos Santos)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schagen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schukraft</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solano Salinas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Souza Filho</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Szynkman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uribe Estrada</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Villavicencio Reyes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walsh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaguna</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zanella Beguelin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analisa Gabriela</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>Roberto</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>Jesus</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Martin</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Alexandre</td>
<td></td>
<td>67</td>
</tr>
<tr>
<td>Danielle</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Eduardo</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>Celio Adrega</td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>Edivaldo</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>Edward</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Gabriela Alejandra</td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>Silvia</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Jonathan</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>Andres</td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>Ana das Merces</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Richard</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>Miguel Angel</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>Andreas</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>Erica</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Juan</td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>Moises</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Marcos</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Veronica</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>J.-Alexis</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>Simon</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Sergio</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>Esteban</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>Federico Andres</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Helena de Fatima</td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Sven</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>Jurgen</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>Carlos Javier</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Luiz Vitor</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Alejandro Andres</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>Cecilia</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Cristian</td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>Roberval</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>Carlos</td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Juan Francisco</td>
<td></td>
<td>66</td>
</tr>
</tbody>
</table>
Contents

Preface ................................................................. v
Photo ........................................................................ vii
List of people in photograph ......................................... ix

FIELD THEORY FOR THE STANDARD MODEL
Ronald Kleiss .......................................................... 1

INTRODUCTION TO QCD
Michelangelo L. Mangano ............................................. 41

BEYOND THE STANDARD MODEL
Esteban Roulet ........................................................ 85

INTRODUCTION TO NEUTRINO PHYSICS
Paolo Lipari ............................................................. 115

LITTLE BANG AT BIG ACCELERATORS:
HEAVY ION PHYSICS FROM AGS TO LHC
Jurgen Schukraft ...................................................... 201

INSTRUMENTATION
Olav Ullaland .......................................................... 217

TRIGGER AND DATA ACQUISITION
Nick Ellis ................................................................ 253
International Scientific Committee ............................. 277
Local Organizing Committee ....................................... 277
Lecturers .................................................................. 277
Discussion Leaders ................................................... 277
List of Students ....................................................... 278
List of Posters ........................................................ 281
Photographs ............................................................ 283
1. THE ULTIMATE SLIT EXPERIMENT: QUANTUM MECHANICS AND THE PATH INTEGRAL

The aim of these lectures is to show how the phenomenology of elementary particles and their interactions can be described in terms of Feynman diagrams. The treatment is not rigorous: rather, I intend to give plausibility arguments, always sticking as closely as possible to physics. In my opinion, the path integral formulation of relativistic quantum physics is one of the most directly physical ways of talking about the phenomena of the elementary particle world. It may therefore be useful to see how the path integral view naturally arises from basic quantum mechanics, using the arguments of Feynman himself.

The primordial quantum-mechanical situation is, of course, the two-slit experiment.

A ‘particle’ (electron, photon, human volunteer, . . .) is emitted at position A at time \( t_1 \), and after a while, at time \( t_2 \), it is observed to be in the ‘detector’, at position B out of all possible positions. Between A and B we have put a screen with 2 holes through which the particle may pass. Quantum mechanics tells us that the real, positive probability of finding the particle at B can be computed as the absolute value squared of the complex-valued probability amplitude, which in this case consists of two contributions, one from each of the two possible paths the particle ‘could have taken’ (indicated by straight lines). The superposition of two complex contributions gives rise to the well-known interference patterns that are observed, demonstrating the wavelike nature of matter. We now become enthusiastic, and put in yet another screen:

Now there are 4 different paths, each giving their contribution, and the interference pattern becomes correspondingly more complicated. Having become fanatical, we now put in many screens, each with many holes:
There are now very many possible paths\(^1\), and we have just drawn a few: all the possible paths give a contribution to the probability amplitude of arriving at B. By now being rabidly frenzied, we decide on the ultimate multislit experiment: we decide to insert an infinite number of screens, each with so many holes in it that every molecule in every screen is completely drilled away. This is simpler than it sounds: the particle now travels through empty space! Still, to be fair we have to allow for all possible paths, three of which are drawn.

Our conclusion must be that all possible paths between A and B give a contribution to the probability amplitude, and we have to sum these over all paths: the path integral.

In many cases, we do not know that the particle was actually at A at time \( t_1 \), and the only information we have is a quantum-mechanical wave-function \( \psi(x_1, t_1) \) that gives us the probability amplitude for finding the particle at a given position \( x_1 \) at time \( t_1 \); by the path integral this is then translated into the wave-function \( \psi(x_2, t_2) \) at the later time \( t_2 \).

But what is the contribution from each path? The correct choice is to take \( \exp(iS/\hbar) \), where \( S \) is the action associated with the path, that is, the time integral over the Lagrangian from \( t_1 \) to \( t_2 \). To see this, let us write the prescription as given, for this one-dimensional case:

\[
\psi(x_2, t_2) = N \int Dx \exp \left[ \frac{i}{\hbar} \int_{t_1}^{t_2} dt \left( \frac{m}{2} \dot{x}^2 - V(x) \right) \right] \psi(x_1, t_1), \tag{1}
\]

where \( x = x(t) \) is any path in time with \( x(t_1) = x_1 \) and \( x(t_2) = x_2 \), and \( Dx \) denotes the super-infinitely large sum over paths. \( m \) is the particle’s mass, and \( V \) any potential in which it moves. The factor \( N \) is a normalization, necessary to ensure the conservation of probability, since \( |\psi|^2 \) must integrate to 1 at all times. The dot denotes a time derivative. Let us now take \( t_2 \) to be infinitesimally later than \( t_1: t_2 = t_1 + \tau \), with \( \tau \) very very small. It is now a good approximation to replace the time integral by a multiplication by \( \tau \). Writing \( x_1 = x_2 + y \), we may then replace Eq. (1) by

\[
\psi(x_2, t_1 + \tau) = N \int dy \exp \left[ \frac{i\tau}{\hbar} \left( \frac{my^2}{2\tau^2} - V(x) \right) \right] \psi(x_2 + y, t_1). \tag{2}
\]

It is easy to see that, if \( my^2/\hbar\tau \) becomes large, the contributions will interfere utterly destructively, so that we may take \( O(y) = O(\tau^{1/2}) \). This allows us to make an expansion in \( y \) and \( \tau \) up to first order in \( \tau \):

\[
\psi(x_2, t_1) + \tau \psi(x_2, t_1) = N \int dy \exp \left( \frac{imy^2}{2\hbar\tau} \right) \times \left( \left( 1 - \frac{i\tau}{\hbar} V(x_2) \right) \psi(x_2, t_1) + y\psi'(x_2, t_1) + \frac{1}{2} y^2\psi''(x_2, t_1) \right), \tag{3}
\]

where primes denote space derivatives, and we have assumed that \( \psi(x, t) \) is smooth enough.

---

\(^1\)The number of paths is the product of the number of screens and the number of holes per screen, even if we forbid backwards motion of the particle.
The zeroth-order term gives us \( N \)

\[
N \int dy \exp \left( \frac{imy^2}{2\hbar} \right) = N \sqrt{2i\pi\hbar/m} = 1, 
\]

from which

\[
N \int dy \ y \exp \left( \frac{imy^2}{2\hbar} \right) = 0, 
\]

\[
N \int dy \ y^2 \exp \left( \frac{imy^2}{2\hbar} \right) = i\hbar/m, 
\]

and the term of first order in \( \tau \) then results in

\[
i\hbar\psi(x,t) = -\frac{\hbar^2}{2m}\psi''(x,t) + V(x)\psi(x,t) 
\]

the Schrödinger equation.

Note that we have not ‘proven’ the Schrödinger equation here: we merely have replaced it, as a basic postulate of quantum physics, by another one, namely the path integral. But the above indicates that the path-integral picture is a useful way of looking at quantum physics. The various paths \( x(t) \) are ‘random’ objects, and the only information we have about them is the probability amplitude connected with them. It is this probabilistic interpretation that we want to take with us. To make the envisaging easier, we shall for now drop the \( i \), and talk about real, positive probabilities rather than complex probability amplitudes.

2. SMALL BEGINNINGS: FIELD THEORY IN 0 + 0 DIMENSIONS

It is often useful to start simply. Therefore, in this Section we shall take a particularly simple model of the universe: we shall take it to consist of one single point. In this universe is a single particle species without any properties such as spin, etc. The ‘field’ describing the particle is a function over space(time), and therefore consists, in this case, of a single value.

2.1. The path integral and Green’s functions

2.1.1. Free theory and sources

In the spirit of the path integral picture, we assume the value \( \varphi \) of the field to be a random number, whose probability density is proportional to \( \exp(-S(\varphi)) \), for some function \( S(\varphi) \), called the action of the model. One of the simplest choices is to assume the density to be a Gaussian:

\[
S(\varphi) = \frac{1}{2}\mu \varphi^2, 
\]

with some constant \( \mu \). We call this the free theory. Since \( \varphi \) is random, all the information that one can possibly compute about it\(^2\) is contained in the moments of the distribution, which are called the Green’s functions of the theory:

\[
G_p \equiv \langle \varphi^p \rangle = N \int d\varphi \ \varphi^p \ e^{-S(\varphi)}, 
\]

\(^2\)You could imagine instead measuring the value that \( \varphi \) actually takes, but that would be experimental rather than theoretical physics.
where the brackets denote the average under $e^{-S(\varphi)}$, and $N$ is the normalization, chosen such that $\langle \varphi^0 \rangle = 1$. Unless specified otherwise, all integrals run from $-\infty$ to $+\infty$. Trivially, we have in this theory

$$G_2 = \frac{1}{\mu}, \quad G_{2k} = \frac{(2k)!}{2^k k!} G_2^k, \quad G_{2k+1} = 0, \quad k \geq 1.$$  \hspace{1cm} (9)

We can combine the information on the Green’s functions $G_p$ into one generating function, as follows:

$$Z(J) = \sum_{p \geq 0} G_p \frac{J^p}{p!} = N \int d\varphi \; e^{-(S(\varphi) - J\varphi)}, \quad G_p = \frac{\partial^p}{\partial J^p} Z(J) \bigg|_{J=0}.$$  \hspace{1cm} (10)

The ‘counting’ number $J$ is called a source; its physical meaning will become clear later on. We can compute $Z(J)$ directly by ‘completing the square’:

$$S(\varphi)^2 - J \varphi = \frac{m u}{2} \left( \varphi - \frac{J}{\mu} \right)^2 - \frac{J^2}{2\mu}$$  \hspace{1cm} (11)

and then doing the Gaussian integral, or by realizing, from its integral representation, that it obeys the differential equation

$$\mu \frac{\partial}{\partial J} Z(J) - J Z(J) = N \int d\varphi \; (\mu \varphi - J) e^{-S(\varphi) + J \varphi}$$

$$= -N \int d\varphi \frac{\partial}{\partial \varphi} e^{-S(\varphi) + J \varphi} = 0,$$  \hspace{1cm} (12)

since we have here a total derivative. Together with $Z(0) = 1$, in either case we arrive at

$$Z(J) = \exp \left( \frac{1}{2\mu} J^2 \right).$$  \hspace{1cm} (13)

2.1.2. Interacting theories and perturbation theory

In real life, we expect theories to be more complicated than just the free one, so let us add a term to $S(\varphi)$ with a higher power of $\varphi$. The simplest choice that still leads to a finite path integral is

$$S(\varphi) = \frac{1}{2} \mu \varphi^2 + \frac{1}{24} \lambda \varphi^4.$$  \hspace{1cm} (14)

Extra terms added to a free theory are called interaction terms, and coefficients like $\lambda$ are called coupling constants\(^3\). Computing the $G_p$ is now immediately much more difficult. A possible approach is to consider the extra term as a small ‘perturbation’ of the Gaussian (especially if $\lambda$ is small), and make an expansion in powers of $\lambda$. For odd $q$, $G_q$ is again zero, while for even $q = 2p$ we have\(^4\)

$$G_{2p} = N \int d\varphi \; \varphi^{2p} e^{-(\frac{1}{2}\mu \varphi^2 + \frac{1}{24} \lambda \varphi^4)}$$

$$= \frac{N}{\mu^p} \sum_{k \geq 0} \left( -\frac{\lambda}{24} \right)^k \frac{(2p+4k)!}{2^{2k} k!(p+k)!}.$$

\(^3\)The factor $1/24$ is a convention: in general, we put a factor $1/q!$ before an interaction of type $\varphi^q$. Some authors use a different convention, so beware.

\(^4\)Here we use the fact that for a Gaussian with zero mean and unit standard deviation, the $2n^{th}$ moment is given by $(2n)!/n!2^n$.
The normalization $N$ is again fixed by $G_0 \equiv 1$,

$$N = \frac{1}{\sqrt{2\pi}} \left( 1 + \frac{1}{8} \frac{\lambda}{\mu^2} - \frac{29}{384} \frac{\lambda^2}{\mu^4} + \frac{107}{1024} \frac{\lambda^3}{\mu^6} + O\left(\lambda^4\right) \right),$$

while the first few nonzero Green’s functions read

$$G_2 = \frac{1}{\mu} - \frac{1}{2} \frac{\lambda}{\mu^3} + \frac{2}{3} \frac{\lambda^2}{\mu^5} - \frac{11}{8} \frac{\lambda^3}{\mu^7} + O\left(\lambda^4\right),$$
$$G_4 = \frac{3}{\mu} - \frac{1}{4} \frac{\lambda}{\mu^4} + \frac{33}{4} \frac{\lambda^2}{\mu^6} - \frac{68}{3} \frac{\lambda^3}{\mu^8} + O\left(\lambda^4\right),$$
$$G_6 = \frac{15}{\mu^3} - \frac{75}{2} \frac{\lambda}{\mu^5} + \frac{445}{4} \frac{\lambda^2}{\mu^7} - \frac{1585}{4} \frac{\lambda^3}{\mu^9} + O\left(\lambda^4\right).$$

This approach is called the perturbation expansion: the effects of interaction are implicitly considered as ‘small corrections’ to the free theory. Clearly, similar but more complicated expressions can be obtained in the same way for other interaction terms, even if the path integral itself is not defined, for instance in a $\varphi^3$ theory (in which case also the odd Green’s functions would be non-zero). The alternative approach would be to use again the integral representation, and now derive another differential equation for $Z(J)$:

$$\frac{1}{6} \lambda \frac{\partial^3}{\partial J^3} Z(J) + \mu \frac{\partial}{\partial J} Z(J) - J Z(J) = 0.$$  

There are 3 independent solutions, none particularly simple; and then we would have to pick the correct one, that gives the perturbative expansion above.

### 2.1.3. Connected Green’s functions

The information about the probability density encoded in the moments $G_p$ can of course also be described in other ways. One of the more useful descriptions (also employed in statistics) is that in terms of the cumulants $C_n$, defined by

$$Z(J) = \sum_{p \geq 0} \frac{1}{p!} J^p G_p = e^{W(J)}, \quad W(J) = \sum_{n \geq 1} \frac{1}{n!} J^n C_n,$$

so that

$$C_n = \left\{ \frac{\partial^n}{\partial J^n} W(J) \right\}_{J=0} = \left\{ \frac{\partial^n}{\partial J^n} \log Z(J) \right\}_{J=0}.$$  

In the jargon, the cumulants are called the connected Green’s functions: the ‘connected’ will become clear later on. From the definition of $Z$ and $W$, it can easily be checked that $C_1 = G_1$ is the mean of the probability density, while $C_2 = G_2 - G_1^2$ is its variance, $C_3$ its skewness. $C_4$ is related to the kurtosis, and so on. It will turn out that these correspond to the physically more interesting properties of the theory. From Eq. (13) we see that for a free theory only $C_2$ is nonzero, which is why it is called a free theory. For the $\varphi^4$ theory we have

$$C_2 = \frac{1}{\mu} - \frac{1}{2} \frac{\lambda}{\mu^3} + \frac{2}{3} \frac{\lambda^2}{\mu^5} - \frac{11}{8} \frac{\lambda^3}{\mu^7} + O\left(\lambda^4\right),$$
$$C_4 = \frac{\lambda}{\mu^4} + \frac{7}{2} \frac{\lambda^2}{\mu^6} - \frac{189}{12} \frac{\lambda^3}{\mu^8} + O\left(\lambda^4\right),$$
$$C_6 = \frac{10 \lambda^2}{\mu^7} - \frac{80 \lambda^3}{\mu^9} + O\left(\lambda^4\right).$$

See, however, the Appendix to this Section.
2.2. Feynman diagrams

The Feynman diagram approach is just another way to arrive at the expansions for $G_p$ and $C_n$. For now, we just give a recipe to get $C_n$: later on we prove that it is indeed the correct one. To obtain $C_n$ ($n \geq 1$) to order $\lambda^k$, take $k$ vertices with 4 legs and join them by lines, in such a way that precisely $n$ lines are sticking out: do this in all possible ways such that the graph is connected, so that you can walk from any external line to any other one over the graph. Closed loops are perfectly allowed, and in fact as $k$ increases become unavoidable. Every such Feynman graph corresponds to a number, as follows: every line carries a factor $1/\mu$, and every vertex a factor $-\lambda$. There is also a numerical factor $1/(n!k!(4!)^k)$, the coefficient of $(-\lambda)^k J^n$ in the expansion of $e^{-S+J\varphi}$, since that is the integral we are actually trying to do. Now add the results of all Feynman graphs. The assignment of numbers to elements of Feynman graphs are called the Feynman rules, and are an important issue in these lectures: the choice of Feynman rules defines the theory. In this case, they are:

$$\begin{align*}
\hline & \rightarrow \frac{1}{\mu}, \\
\times & \rightarrow -\lambda .
\end{align*}$$

Of course, there are many diagrams that look exactly the same, differing only in the use of one of the $k$ vertices rather than another, or one of its legs rather than another, while the diagram is being built up. It is customary to collect all such topologically equivalent diagrams into one. The numerical factor in front is then $1/f$, where $f$ is the number of ways in which legs and vertices can be interchanged without changing the diagram. This is called the order of the symmetry group of the diagram, and $1/f$ is called its symmetry factor. Some examples of diagrams with their symmetry factors are:

$$\begin{align*}
\hline & \rightarrow 1 , \\
\times & \rightarrow \frac{1}{2} , \\
\bigotimes & \rightarrow \frac{1}{6} , \\
\bigcirc & \rightarrow \frac{1}{12} .
\end{align*}$$

Note that the external legs are counted as distinct. The determination of symmetry factors, although in principle straightforward, can be very cumbersome; fortunately in many cases they are simple or even (as in QED, as we shall see) always unity. As an exercises, you may check that

$$C_2 = \hline + \times + \bigotimes + \bigotimes + \bigcirc + O(\lambda^3) .$$

It should be kept in mind that the Feynman diagrams are just funny ways of writing definite mathematical objects: it makes perfect sense to talk about the numerical value of a diagrams.

2.2.1. The Schwinger-Dyson equation

The drawing of Feynman diagrams can be systematized. To this end, we introduce the Feynman rule for the source $J$, which we now picture by another vertex (denoted by a small cross):

$$\begin{align*}
\hline & \rightarrow +J ,
\end{align*}$$

the $+$ sign coming from the fact that we added $-J\varphi$ to the action in Eq. (10). The physical meaning of the source is now apparent: a source acts as an object that gives rise to a line entering the diagram from the ‘outside’. In real life, a source may be also described by Feynman diagrams (as in QED, where an

\[\text{It is also one of the things that is very difficult to program in a computer.}\]
electron line may emit a photon, and acts as a source for this photon), or we may assume the source to be `infinitely far away’ so that the line comes in from `infinity’, as it should for, say, electrons and positron coming out of an accelerator into the interaction region (we shall come back to this point later on). The symmetry factor will now also contain $1/k!$ in diagrams where $k$ such source vertices are present. Each derivative with respect to $J$ corresponds to deleting one such vertex, so that an extra external line is generated. Let us now consider the set of all possible connected diagrams with precisely one external leg. Since (the value of) this set depends on $J$, we denote it by $\phi(J)$ (beware: $\varphi$ stands for the `dummy’ integration variable in the path integral, and $\phi$ denotes a definite number, namely the value of the set of diagrams. There is a good reason for this confusing notation, as we shall see below). Denoting the set of all connected diagrams by a shaded blob, we have

$$
\begin{align*}
\text{blob} &= \phi(J), \\
\text{blob} &= \frac{\partial}{\partial J} \phi(J), \\
\text{blob} &= \frac{\partial^2}{\partial J^2} \phi(J), \cdots
\end{align*}
$$

(note that this works because we put in a factor $1/n!$ with $J^n$, as dictated by our considerations on the symmetry factors), and also

$$
C_{n+1} = \left[ \frac{\partial^n}{\partial J^n} \phi(J) \right]_{J=0}.
$$

Let us now follow the external line of $\phi(J)$ into the diagram. There are several possibilities. The line may immediately encounter a $J$ vertex, and stop there, or a $\lambda$ vertex, and split into three. These lines may not come together again, or maybe two of them will, or all three. Diagrammatically, this reads

$$
\begin{align*}
\text{blob} &= \times + \frac{1}{6} \quad \text{blob} + \frac{1}{2} \quad \text{blob} + \frac{1}{6} \quad \text{blob},
\end{align*}
$$

where we have to put in the symmetry factors corresponding to the interchange of equivalent internal lines. This is the Schwinger-Dyson equation for this theory: in terms of $\phi(J)$, it reads

$$
\phi(J) = \frac{J}{\mu} - \frac{\lambda}{6\mu} \phi(J)^3 - \frac{\lambda}{2\mu} \phi(J) \frac{\partial}{\partial J} \phi(J) - \frac{\lambda}{6\mu} \frac{\partial^2}{\partial J^2} \phi(J).
$$

The Schwinger-Dyson (SD) equation allows us to construct the whole set of connected diagrams in a recursive manner. That we get the right diagrams can simply be proven: if we are correct, then we should have

$$
\phi(J) = \frac{\partial}{\partial J} W(J) = \frac{1}{Z(J)} \left( \frac{\partial}{\partial J} Z(J) \right).
$$

By making this substitution for $\phi(J)$ in Eq. (18) we indeed obtain Eq. (26); and since the simplest term in $\phi(J)$, namely $J/\mu$, is correct, all the other ones are correct as well. Other theories have their own SD equation: if the action is

$$
S(\varphi) = \frac{1}{2} \mu \varphi^2 + \sum_{r \geq 3} \frac{\lambda_r}{r!} \varphi^r,
$$

---

7 Lines attached to a $J$ vertex count as internal lines.

8 Note that this is only unambiguous if the blobs denote connected diagrams.

9 Indeed, it can easily be implemented in computer algebra, starting with $\phi(J) = J/\mu$ and iterating the right-hand side of Eq. (26).
then the SD equation in terms of \( Z(J) \) reads
\[
\mu \frac{\partial}{\partial J} Z(J) + \sum_{r \geq 3} \frac{\lambda_r}{(r-1)!} \frac{\partial^{r-1}}{\partial J^{r-1}} Z(J) = S' \left( \frac{\partial}{\partial J} \right) Z(J) = J Z(J) .
\] (29)

It is left as an exercises to work out the diagrammatic form for a theory with \( \varphi^3 \) as well as \( \varphi^4 \) interactions. It must be remarked that, once the external leg that serves as the ‘starting point’ has been chosen, we can for every diagram unambiguously determine to which term in the SD equation it belongs, just by looking at what happens at the ‘first vertex’. Therefore, the symmetry factor of the diagram is completely determined by the factors \( \frac{1}{2} = 2 \) and \( \frac{1}{6} = 6 \) in the SD equation\(^{10}\). In a theory like QED, where there is only one vertex, where three non-equivalent lines meet, all symmetry factors are therefore unity. The same holds for the electroweak model as long as no four-boson vertices or Higgs self-interactions are involved.

Finally, it may be realized that \( \phi(J) \) has a simple interpretation: in the presence of the source \( J \), the random variable \( \varphi \) has probability density \( N \exp(-S(\varphi) + J \varphi) \) rather than \( N \exp(-S(\varphi)) \). Denoting by \( \langle \cdots \rangle_J \) averages with respect to this density, we see that
\[
\phi(J) = \langle \varphi \rangle_J ,
\] (30)
which explains our use of two such similar symbols.

2.2.2. Connected, disconnected and vacuum diagrams

Having seen that the set of all connected diagrams without external legs but with legs ending in \( J \) vertices represents \( W(J) \), we now understand where the name ‘connected Green’s function’ comes from. The set of Green’s functions in \( Z(J) \) is then
\[ e^{W(J)} = 1 + W(J) + W(J)^2/2 + \cdots , \] and we see that it corresponds to the set of all diagrams, both connected ones and disconnected ones (that consist of two or more separate connected pieces), with an extra symmetry factor \( 1/m! \) for each connected diagram that occurs precisely \( m \) times as a factor in a given term in \( Z(J) \). A disconnected diagram consisting of \( p \) connected pieces comes from the term \( W(J)^p/p! \) in \( e^{W(J)} \). Only one type seems to be missing, namely those diagrams that do not contain any \( J \) vertex at all: this set is given by
\[
\exp \left( \infty + \infty + \infty + \cdots \right) = \exp \left( -\frac{1}{8} \frac{\lambda}{\mu^2} + \frac{1}{12} \frac{\lambda^2}{\mu^4} + \cdots \right) .
\] (31)

That this is not really an omission becomes clear if we realize that these are just the diagrams that crop up in the computation of \( Z(0) \), and hence are always precisely absorbed in the definition of the normalization \( N \): since the same set of vacuum diagrams always occurs as a factor in any \( G_p \), it is always divided out.

2.2.3. The loop expansion and classical theory

The perturbative expansion in \( \lambda \) is straightforward; but in a theory with more than one coupling constant ambiguities may arise, since \textit{a priori} the relative orders of these constants are not known: if the action is
\[
S(\varphi) = \frac{1}{2} \mu \varphi^2 + \frac{1}{6} \lambda_3 \varphi^3 + \frac{1}{24} \lambda_4 \varphi^4 ,
\] (32)
then are we to take \( \lambda_3 \) and \( \lambda_4 \) as being of the same order? Or should we take, say \( \lambda_3^2 \) to be of order \( \lambda_4 \)? There is a more systematic way of ordering the diagrams in a given \( C_n \), by the number of closed loops.

\(^{10}\)It is not, however, a simple product of such factors, since for individual diagrams the lines entering the loops may be inequivalent: the equivalence holds only for the set of all diagrams.
in each diagram. Let us decide to assign a factor $\hbar$ to each loop\textsuperscript{11}. In the SD equation, we can simply account for it by modifying Eq. (25):

$$\phi(J) = \frac{J}{\mu} - \frac{\lambda}{6\mu} \phi(J)^3 - \frac{h\lambda}{2\mu} \phi(J) \frac{\partial}{\partial J} \phi(J) - \frac{h^2 \lambda}{6\mu} \frac{\partial^2}{\partial J^2} \phi(J) . \quad (34)$$

This ordering of the diagrams is called the loop expansion. To account for the occurrence of $\hbar$, we have to modify our definitions a bit: from here on, we define

$$\phi(J) = \hbar \frac{\partial}{\partial J} \log Z(J) , \quad Z(J) = N \int d\varphi \exp \left( -\frac{1}{\hbar} (S(\varphi) - J\varphi) \right) \quad (35)$$

and the SD equation for $Z(J)$ now becomes

$$S'(\hbar \frac{\partial}{\partial J}) Z(J) = JZ(J) . \quad (36)$$

The relative order of $\lambda_3$ and $\lambda_4$, say, can now be fixed: by replacing, in the new definition of $Z(J)$, the integration variable $\varphi$ by $\varphi' = \varphi \sqrt{\hbar}$, we see that $\mu, \lambda_3 \sqrt{\hbar}$, and $\lambda_4 \hbar$ are all of the same order, and therefore $\lambda_3^2$ has the same order as $\lambda_4$ in the loop expansion.

It may be interesting to see what happens when $\hbar$ becomes vanishingly small: this is called the classical limit. On the one hand, we can simply put $\hbar = 0$ in Eq. (34): for a general theory, this gives

$$\phi_c(J) = \frac{J}{\mu} - \sum_{r \geq 3} \frac{\lambda_r}{(r - 1)! \mu} \phi_c(J)^{r-1} \Rightarrow$$

$$\mu \phi_c(J) + \sum_{r \geq 3} \frac{\lambda_r}{(r - 1)!} \phi_c(J)^{r-1} = S'(\phi_c(J)) = J , \quad (37)$$

from the definition of $S(\varphi)$ as a series in $\varphi$. Since $\phi_c(J)$ contains only diagrams without closed loops, it is also called the tree approximation. In the path integral, on the other hand, the integrand will show extremely narrow peaks as $\hbar \to 0$, so that $\phi(J) = \langle \varphi \rangle_J$ will really only get contributions from those positions where the exponent has maxima: these correspond, of course, precisely to those values of $\varphi$ where $S(\varphi) - J\varphi$ has a minimum on the integration axis. The lowest of these minima (assuming that it is unique) gives the dominant contribution, and for this value, $\varphi_c$, we of course have

$$S'(\varphi_c) = J . \quad (38)$$

This is called the classical field equation. Note that the other minima will give contributions that are suppressed relative to the dominant one by $e^{-n/\hbar}$, with $n$ some positive number: such contributions, that are of course totally invisible in an expansion in powers of $\hbar$, are inherently non-perturbative. Usually one tries to arrange things such that $\varphi_c = 0$ when $J$ vanishes\textsuperscript{12}; the other solutions of the classical field equation are called instantons.

\textsuperscript{11}Here, $\hbar$ is of course just an arbitrary numerical constant: its significance as a constant of nature becomes only clear in the more-dimensional case.

\textsuperscript{12}This may necessitate a shift in the definition of $\varphi$, as in the case of spontaneous symmetry breaking.
2.2.4. The effective action

Since the classical equation looks simpler than the full quantum-mechanical one, the question arises: can we find another action (called the effective action) such that its classical approximation coincides with \( \phi(J) \)? If so, we would need to compute only tree diagrams, which is certainly simpler than also computing loops. Denoting the effective action by \( \Gamma(\varphi) \), it must therefore be such that

\[
\Gamma(\phi(J)) = J .
\]

Let us assume that the relation between \( J \) and \( \phi \) can be inverted, at least in some neighbourhood of \( J = 0 \):

\[
\phi = \phi(J) \quad \leftrightarrow \quad J = y(\phi) .
\]

Integration by parts, and using \( phi \, dJ = dW \), then gives us \( \Gamma \):

\[
\Gamma(\varphi) = \int y(\phi) \, d\phi = \phi y(\phi) - \hbar W(y(\phi)) .
\]

The effective action is the Legendre transform of \( W(J) \).

In terms of Feynman diagrams, it is useful to introduce the notion of a one-particle irreducible (1PI) diagram. Such a diagram cannot be made disconnected by cutting through one internal line. For example:

\[
\text{is 1PI} \quad ; \quad \text{is not 1PI} .
\]

A single vertex also counts as 1PI, since there aren’t any internal lines to cut. Let us now redo the reasoning that led to the SD equation. Coming into the diagram, we shall encounter a 1PI piece of it – possibly only a single vertex. Out of this piece stick 0, 1, 2, 3, \ldots lines that, when cut, do make the diagram fall apart. Denoting the 1PI pieces with dark blobs, we therefore have

\[
\begin{align*}
\Gamma(\varphi) &= \gamma_1 \varphi + \frac{1}{2} \mu^2 \varphi^2 + \frac{1}{2} \lambda^2 \varphi^4 + \sum_{r \geq 3} \frac{\gamma_r}{r!} \varphi^r \cdot \\
&= \mu \phi(J) + \sum_{r \geq 1} \frac{\gamma_r}{(r-1)!} \phi(J)^{r-1} = J ,
\end{align*}
\]

where \( -\gamma_r \) denotes the sum of all 1PI diagrams with at least one vertex and precisely \( r \) external legs. Comparing with Eq. (37), we see that the effective action is therefore given by

\[
\Gamma(\varphi) = \gamma_1 \varphi + \frac{1}{2} (\mu + \gamma_2) \varphi^2 + \sum_{r \geq 3} \frac{\gamma_r}{r!} \varphi^r .
\]

Note that since it is usually easy to construct 1PI diagrams with very many external legs, the sum over \( r \) runs up to infinity. For the pure \( \varphi^4 \) theory, we obtain\(^{13}\)

\[
\Gamma(\varphi) = \left( \frac{1}{2} \mu \varphi^2 + \frac{1}{24} \lambda \varphi^4 \right) + \frac{\hbar}{2} \log \left( 1 + \frac{\lambda \varphi^2}{2 \mu} \right) - \frac{\hbar^2 \lambda^2 \varphi^2}{24 \mu^2} \left( \frac{40 \mu^2 + 18 \lambda \varphi^2 + 3 \lambda^2 \varphi^4}{(2 \mu + \lambda \varphi^2)^3} \right) + O(\hbar^3) .
\]

\(^{13}\)By first rewriting Eq. (34) as an equation for \( y(\varphi) \) rather than for \( \phi(J) \), then solving it perturbatively as a power series in \( \hbar \), and then integrating over \( \varphi \).
2.2.5. Dyson summation

In the foregoing, it might seem that the $\varphi^2$ term in the action is treated differently from the interaction terms: after all, the quadratic term occurs like $1/\mu$ in the Feynman rules, while the coupling constants appear with positive power. That this is not really the case becomes clear when we consider the action

$$S(\varphi) = \frac{1}{2} \mu \varphi^2 + \frac{\lambda_2}{2} \varphi^2 + \sum_{r \geq 3} \frac{\lambda_r}{r!} \varphi^r$$

and decide to treat the $\lambda_2$ as a perturbation. We then have

$$\varnothing = \mathbf{x} + \mathbf{\bullet} + \text{(higher interaction terms)},$$

where the fat dot stands for the $\lambda_2$ vertex, or

$$\phi(J) = \frac{J}{\mu} - \frac{\lambda_2}{\mu} \phi(J) - \frac{1}{\mu} \text{(higher interaction terms)},$$

and we have indicated all occurrences of $\mu$. By bringing the term linear in $\phi(J)$ to the left, we can see that this is exactly the same as

$$\phi(J) = \frac{J}{\mu + \lambda_2} - \frac{1}{\mu + \lambda_2} \text{(higher interaction terms)},$$

in other words, we might as well have absorbed the $\lambda_2$ into the $\mu$ from the start. This absorption of quadratic interactions into the $\mu$ term is called Dyson summation. The apparently different treatment of quadratic terms is related to the fact that the perturbation expansion is one around the Gaussian form in the path integral.

2.3. Renormalization

So far, we have treated all parameters in the action as if they were in some way logically prior to the physics they imply. In reality, of course, it is just the other way around: in the practice of physics, one tries to obtain the couplings etc. from some measurement experiment by a fit to data, and then uses them in a prediction for the result of some other measurement. It is only the measurements that have a claim to physical reality: strictly speaking, masses and couplings are just bookkeeping devices. This is at the basis of renormalization: if, by some effort, we have computed yet one higher order (in $\hbar$, say) in our prediction, this is useless unless we have the corresponding higher precision in our extraction of the parameters from their measurement. It also (and fortunately!) means that if two parameters always occur in precisely the same combination in our expressions for the measurement and prediction, we cannot, and should not try to, disentangle them. This is the way in which the infamous loop divergences in quantum field theory become ‘absorbed’ into the physical parameters, but we must keep in mind that, even for a perfectly finite theory, renormalization is always necessary as long as we are working in perturbation theory.

2.3.1. Finite renormalization in zero dimensions

To illustrate the above, let us again consider $\varphi^4$ theory. If we include the correct factors $\hbar$, the first few connected Green’s functions are given by

$$C_2 = \frac{1}{\mu} - \frac{\hbar \lambda}{2 \mu^3} + \frac{2 \hbar^2 \lambda^2}{3 \mu^5} - \frac{11 \hbar^3 \lambda^3}{8 \mu^7} + \mathcal{O}(\hbar^4),$$
\[ C_4 = -\frac{\lambda}{\mu^3} + \frac{7h \, \lambda^2}{2 \, \mu^6} - \frac{189h^2 \, \lambda^3}{12 \, \mu^8} + \mathcal{O}\left( h^3 \right) , \]
\[ C_6 = 10 \frac{\lambda^2}{\mu^7} - 80h \frac{\lambda^3}{\mu^7} + \mathcal{O}\left( h^2 \right) . \]  
(50)

There are 2 parameters, \( \mu \) and \( \lambda \), so we need 2 measurements to fix them. Let us take these to be the measurements of \( C_2 \) and \( C_4 \). From the measured values of \( C_2 \) and \( C_4 \), we extract our parameters and then use them to predict \( C_6 \), say. Each new order in \( h \) will change our extracted value of \( \mu \) and \( \lambda \), and so we should write
\[ \mu = \mu_0 + \mu_1 h + \mu_2 h^2 + \mu_3 h^3 + \cdots , \]
\[ \lambda = \lambda_0 + \lambda_1 h + \lambda_2 h^2 + \lambda_3 h^3 + \cdots . \]  
(51)

At each fixed order in perturbation theory, we can extract the truncated values:
\[ \mu^{(p)} = \sum_{k=0}^{p} \mu_k h^k , \quad \lambda^{(p)} = \sum_{k=0}^{p} \lambda_k h^k . \]  
(52)

These lead, to the correct order in \( h \), to similarly truncated approximations to \( C_6 \), which we call \( C_6^{(p)} \). At the lowest (tree) order, we simply have
\[ \mu^{(0)} = \mu_0 = 1 \frac{C_2}{C_4} , \quad \lambda^{(0)} = \lambda_0 = -\frac{C_4}{C_2} , \quad C_6^{(0)} = 10 \lambda_0^2 \mu_0^7 , \]  
(53)

To first order, we have to determine \( \mu_1 \) and \( \lambda_1 \) such that the values of \( C_2 \) and \( C_4 \) remain unchanged, and update the (truncated) prediction for \( C_6 \):
\[ \mu_1 = -\frac{\lambda_0}{2 \mu_0} , \quad \lambda_1 = \frac{2 \lambda_0^2}{3 \mu_0^2} , \quad C_6^{(1)} = 10 \lambda_0^2 \mu_0^7 \left[ -15h \lambda_0^3 \mu_0^9 \right] , \]  
(54)

and so on. For the sake of numbers, let us assume that \( C_2 = 1 \) and \( C_4 = -2 \), and take \( h = 0.01 \). In the Table we give the updated values of \( \mu \), \( \lambda \) and \( C_6 \).

<table>
<thead>
<tr>
<th>( \mathcal{O}(h^p) )</th>
<th>( \mu^{(p)} )</th>
<th>( \lambda^{(p)} )</th>
<th>( C_6^{(p)} )</th>
<th>wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000000</td>
<td>2.000000</td>
<td>40.000000</td>
<td>40.000000</td>
</tr>
<tr>
<td>1</td>
<td>0.990000</td>
<td>2.060000</td>
<td>38.800000</td>
<td>33.600000</td>
</tr>
<tr>
<td>2</td>
<td>0.989767</td>
<td>2.060600</td>
<td>38.872000</td>
<td>34.41867</td>
</tr>
<tr>
<td>3</td>
<td>0.989766</td>
<td>2.060622</td>
<td>38.866000</td>
<td>34.31619</td>
</tr>
<tr>
<td>4</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86662</td>
<td>34.32951</td>
</tr>
<tr>
<td>5</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86654</td>
<td>34.32767</td>
</tr>
<tr>
<td>6</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86655</td>
<td>34.32795</td>
</tr>
<tr>
<td>7</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86655</td>
<td>34.32790</td>
</tr>
<tr>
<td>8</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86655</td>
<td>34.32791</td>
</tr>
<tr>
<td>9</td>
<td>0.989766</td>
<td>2.060621</td>
<td>38.86655</td>
<td>34.32791</td>
</tr>
</tbody>
</table>

The column labelled ‘wrong’ is the result for \( C_6 \) where we have just included the higher orders in the prediction, while keeping \( \mu \) and \( \lambda \) fixed to their lowest-order values 1 and 2. For this value of \( h \), perturbation theory appears to do nicely – but if you feel happy now, read the Appendix!
2.3.2. Divergences and renormalization: a toy model

In real, four-dimensional life, the higher-order corrections are of course both more complicated than just simple combinations of $\mu$ and $\lambda$, and more dangerous, since loop diagrams tend to give divergences. Fortunately, in many cases these are automatically taken care of by renormalization, provided the theory has the right structure. In order to see how this works, we shall study a simple toy model.

In order to include, in an admittedly crude way, the divergence structure of the loop diagrams in the four-dimensional theory into our zero-dimensional model, let us add a complication to the zero-dimensional Feynman rules, as follows. We decide that every closed loop with precisely one vertex on it shall be multiplied by a factor $1 + c_1$, and every closed loop with precisely 2 vertices on it by a factor $1 + c_2$; loops with more vertices are not modified. Here, $c_1$ and $c_2$ are numbers that eventually go to infinity. To handle this, we assume that they both depend on some parameter $\Lambda$, and diverge as $\Lambda \to \infty$. Diagrammatically, we have the replacement

$$
\begin{align*}
\begin{array}{c}
\text{loop with one vertex} \\
\text{loop with two vertices}
\end{array}
\end{align*}
\to
\begin{align*}
\begin{array}{c}
\text{loop with one vertex} \\
\text{loop with two vertices}
\end{array} + \frac{c_1}{\mu} \text{ and } \frac{c_2}{\mu^2},
\end{align*}
\quad c_1 \times \frac{c_2}{\mu^2}
$$

(55)

Let us consider a theory with both $\varphi^3$ and $\varphi^4$ couplings:

$$
S(\varphi) = \lambda_1 \varphi + \frac{1}{2} \mu \varphi^2 + \frac{1}{6} \lambda_3 \varphi^3 + \frac{1}{24} \lambda_4 \varphi^4.
$$

(56)

Note that we have also introduced a term linear in $\varphi$. This is because we would like to interpret particles as fluctuations in the field: for the situation with no particles, the vacuum, it is most reasonable to also require the field to be absent. Hence, we require that $\langle \varphi(0) \rangle = \langle \varphi \rangle = 0$. For pure $\varphi^4$ theory this is automatically ensured by the $\varphi \leftrightarrow -\varphi$ symmetry of the action, but when a $\varphi^3$ term is present, the coupling $\lambda_1$ has to be renormalized to achieve this.

Of course, the SD equation will have to be modified in such a way that now also the dotted diagrams are automatically generated, symmetry factors and all. This can be done by introducing new vertices $k_1, \ldots, k_4$, as follows:

$$
\begin{align*}
\begin{array}{c}
\text{loop with one vertex} \\
\text{loop with two vertices} \\
\text{loop with three vertices} \\
\text{loop with four vertices}
\end{array}
\equiv
\begin{array}{c}
\frac{-k_1}{\mu} \\
\frac{-k_2}{\mu^2} \\
\frac{-k_3}{\mu^3} \\
\frac{-k_4}{\mu^3}
\end{array}
\times
\begin{array}{c}
\text{loop with one vertex} \\
\text{loop with two vertices} \\
\text{loop with three vertices} \\
\text{loop with four vertices}
\end{array} + \cdots,
\end{align*}
\quad (57)
$$

Writing a dot for the $\lambda_1$ vertex, we can now give the modified SD equation that includes the dotted diagrams correctly:
2.3.3. Nonrenormalizable theories

The above toy model is of course very crude: in reality, there are many more divergent loop diagrams than just the ones we used. These are called overlapping divergences, occurring for instance in the diagram \( \overline{\Box} \), which in our toy model has no dotted partner. The significant point is that the divergences are located in diagrams (or sub-diagrams) with just so many legs that they can be absorbed into a combination with bare vertices of the theory. That this is not always possible can be seen from some examples. Suppose we also had a divergence in loops with 3 vertices. The following diagram would then occur:

\[
\begin{align*}
\chi_1 &= \chi_3 + \chi_4 \\
\mu' &= \mu + k_2 \\
\lambda' &= \lambda + 3k_4 \\
\lambda &= \lambda + 3k_4
\end{align*}
\]

where we have indicated the powers of \( h \) and the symmetry factors coming from equivalent lines (note that in the new ‘box’ vertices not all lines are equivalent). We can now see that if we define

\[
\chi_1 = \chi_3 + \chi_4, \quad \mu' = \mu + k_2, \quad \lambda_3 = \lambda + 3k_3, \quad \lambda_4 = \lambda + 3k_4,
\]

this SD equation is exactly that corresponding to the action

\[
S(\varphi) = \chi_1 \varphi + \frac{1}{2} \mu' \varphi^2 + \frac{1}{6} \lambda_3 \varphi^3 + \frac{1}{24} \lambda_4 \varphi^4.
\]

In other words, the bare parameters \( \mu \) and \( \lambda \) are themselves invisible, and only the renormalized parameters \( \mu' \) and \( \chi_1 \) occur. The principle of renormalization then forces us to tune the bare parameters in such a way that they completely absorb all effects of \( c_{1,2} \). Admittedly this means that the bare parameters become divergent as well – but since they show up nowhere by themselves, this is actually irrelevant.

### 2.3.3. Nonrenormalizable theories

The above toy model is of course very crude: in reality, there are many more divergent loop diagrams than just the ones we used. These are called overlapping divergences, occurring for instance in the diagram \( \overline{\Box} \), which in our toy model has no dotted partner. The significant point is that the divergences are located in diagrams (or sub-diagrams) with just so many legs that they can be absorbed into a combination with bare vertices of the theory. That this is not always possible can be seen from some examples. Suppose we also had a divergence in loops with 3 vertices. The following diagram would then occur:

which would have to be absorbed into a 6-point vertex. In a \( \varphi^4 \) theory there is no such vertex, and the divergence would remain. If we try to save the day by going to a theory with additional \( \varphi^6 \) couplings, then again there would be diagrams like

which would necessitate a \( \varphi^8 \) vertex, and so on. Although a theory with an infinite number of parameters is by itself not forbidden, it would require us to make an infinite number of measurements before we could start on any prediction, and physics would be hopeless. Such theories are called nonrenormalizable.
Suppose the largest power of $\varphi$ in the action is $p$. A loop with up to $k$ vertices on it can then have up to $k(p-2)$ legs sticking out from it. If it is divergent, the theory will only be renormalizable if $k(p-2) \leq p$. If $k = 1$, all theories are renormalizable; for $k = 2$, only $p \leq 4$ is allowed, and this is the reason why $\varphi^4$ theory is used as our paradigm\textsuperscript{14}.

2.3.4. Running couplings and the $\beta$ function

So far, we have not discussed the $\Lambda$ dependence of the renormalization procedure. Typically, $\Lambda$ contains some parameter that regularizes the loop divergences (like the famous $1/\epsilon$ in dimensional regularization) but it may also contain physical information. In actual theories like QCD, this can for instance be the energy scale at which the measurements ($C_2$ and $C_4$ in the above) are performed, but in general it might be anything else, for instance the time of day or the altitude at which the measurements are performed. We shall simply assume that there is some finite parameter $s$, the scale, such that a change in $s$ leads to a change in $\Lambda$. We can always choose $s$ such that the dependence is linear with unit derivative: $d\Lambda = ds$ (note that what we mean here by ‘scale’ is not the same as what is usually implied: in QCD, for instance, our ‘scale’ is $\log(Q^2)$ rather than $Q^2$).

Now suppose that the measurements are performed at two different scales, $s_1$ and $s_2$. The bare parameters that we extract may be divergent (and we need to regularize), but their extracted values should be the same in both cases, since the action doesn’t know about external experimental information like the scale. This means that the renormalized parameters at the two scales should be related, and that having performed the measurement at scale $s_1$ we ought to be able to predict what would have been the outcome if we had performed it at scale $s_2$ instead. The outcome might be different, and in that case we say that the renormalized parameters are scale-dependent; but the very least requirement is that at both scales the renormalized parameters are finite!

In order to illustrate all this let us suppose that we have a theory with a single dimensionless parameter $v$. An example is QCD with massless quarks, where $v = \alpha_s$. After absorbing the loop divergences we obtain a finite renormalized parameter, which we shall call $w$. This depends on the scale via its dependence on $\Lambda$:

$$w(s) = F(v, \Lambda).$$

(61)

For simplicity, we assume that $F(v, 0) = v$ (this is not really drastic: if $F(v, 0) = f(v)$ is not simply $v$, we just take $f^{-1}(F(v, \Lambda))$ and call that the renormalized parameter $w$, simply a finite function of the old one).

Since $v$ can by assumption be extracted from the measurement of $w(s)$, we also have the inverse function:

$$v = G(w, \Lambda), \quad v = G(F(v, \Lambda), \Lambda).$$

(62)

A small change in scale $s \rightarrow s + ds$ or, equivalently, $\Lambda \rightarrow \Lambda + d\Lambda$, should lead to a corresponding change in $w$:

$$\frac{d}{ds}w(s) \equiv \beta(w, \Lambda) = F_2(v, \Lambda) = F_2(G(w, \Lambda), \Lambda),$$

(63)

where a subscript $i$ denotes partial derivation with respect to the $i^{th}$ argument. We have

$$F_1 G_1 = 1, \quad F_2 G_1 + G_2 = 0,$$

(64)

and therefore

$$\beta(w, \Lambda) = \frac{G_2(w, \Lambda)}{G_1(w, \Lambda)}.$$

(65)

\textsuperscript{14}This is also the reason why the self-interactions of the Higgs boson in the Standard Model are chosen to be of type $\varphi^3 + \varphi^4$, and not higher.
If we insist that the renormalized parameter at the new scale is also finite, all occurrence of $\Lambda$ should disappear in the $\beta$ function:

$$\frac{\partial}{\partial \Lambda} \beta(w, \Lambda) = 0 \implies \frac{d}{ds} w(s) = \beta(w) .$$

(66)

The last equation is known as the renormalization group equation. It implies that $G$ can be written in the form\(^{15}\)

$$G(w, \Lambda) = G(-\Lambda + b(w)) , \quad b(w) = \int \frac{1}{\beta(w)} dw .$$

(67)

Since $w = v$ for $\Lambda = 0$, the functions $G$ and $b$ are each other’s inverse. This also means that

$$w = G(\Lambda + b(v)) = F(v, \Lambda) , \quad \frac{F_2(v, \Lambda)}{F_1(v, \Lambda)} = \beta(v) .$$

(68)

Thus, knowledge of $\beta(w)$ completely fixes the form of $F$: if a theory results in an $F$ with a different form, it is only meaningful at precisely the scale at which the measurements are made, but not at any other one.

In perturbation theory, we can compute the first few terms in $F$ by hand, given the beta function:

$$\beta(v) = \beta_0 v^2 + \beta_1 v^3 + \beta_2 v^4 + \cdots ,$$

$$F(v, \Lambda) = v + \beta_0 \Lambda v^2 + \left( \beta_0^2 \Lambda^2 + \beta_1 \Lambda \right) v^3 + \left( \beta_0^3 \Lambda^3 + \frac{5}{2} \beta_0 \beta_1 \Lambda^2 \beta_2 \Lambda \right) v^4 + \cdots .$$

(69)

The condition $F(v, 0) = v$ that we started with is of course just one of the possibilities, depending amongst other things on the precise definition of $\Lambda$. Different such renormalization prescriptions or schemes are possible, and we just used a particularly simple one.

The renormalized parameters in two different schemes must be related in a finite manner: if we adopt another scheme in which the renormalized parameter is $w'$, then we have

$$w = h(w') = w' + h_1 w'^2 + h_2 w'^3 + h_3 w'^4 + \cdots ,$$

(70)

which leads to a new $\beta$ function, $\beta'$:

$$\beta'(w') = \frac{d}{ds} w' = \frac{dw'}{dw} \beta(w) = \frac{1}{h'(w')} \beta(h(w')) ,$$

(71)

where $h'$ denotes the derivative. It is easy to check that the first two coefficients in the expansions of $\beta(w)$ and $\beta'(w')$ coincide: scheme dependence shows up in the third and higher terms.

In theories with more parameters we have a set of coupled renormalization group equations that have to be solved simultaneously.

2.4. Appendix: Convergence of perturbation theory

In setting up perturbation theory we have assumed that in $\varphi^4$ theory, say, the interaction term is a ‘small disturbance’ of the Gaussian shape of the free theory. Close to the peak this may be true, but for large $\varphi$ values it is not: the tails of the distribution are nothing like Gaussian. As a consequence, perturbation theory does not converge! To see this, look again at the $G_{2\rho}$ of Eq. (15), which (putting in the $h$) we can

\(^{15}\)By separation of variable in Eq. (65).
also write as

\[ G_{2p} = N \sqrt{2\pi} \frac{(2p)!}{p!} \left( \frac{h}{2\mu} \right)^{p} \sum_{k \geq 0} (-)^{k} T_{k}, \]

\[ T_{0} = 1, \quad \frac{T_{k}}{T_{k-1}} = \frac{\hbar \lambda (2p + 4k - 1)(2p + 4k - 3)}{24k}. \]  

(72)

The radius of convergence of this series\(^{16}\) is zero for any nonzero \( \lambda \), since \( T_{k}/T_{k-1} \to \infty \) as \( k \to \infty \). For small \( \lambda \), the terms \( T_{k} \) will start out decreasing, but at \( k \sim 3\mu^{2}/2\hbar \lambda \) they will start increasing again, and the perturbative prediction will start to oscillate wildly: perturbation theory breaks down at order \( 3\mu^{2}/2\hbar \lambda \). For \( \mu = 1, \lambda = 2, \hbar = 0.01 \) this means around the 75th order, so it is not surprising that the Table in Section 2.3.1. looked reliable. This breakdown is a general feature as long as we expand the cubic or higher interactions around the quadratic part. That the problem is not cured by either dropping the vacuum diagrams, going over to connected Green’s functions, or applying renormalization, becomes clear when we redo the finite renormalization procedure for \( C_{6} \), this time for \( \hbar = 0.3 \):

<table>
<thead>
<tr>
<th>( \mathcal{O}(\hbar^{p}) )</th>
<th>( \mu^{(p)} )</th>
<th>( \lambda^{(p)} )</th>
<th>( C_{6}^{(p)} )</th>
<th>wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000000</td>
<td>2.000000</td>
<td>40.00000</td>
<td>40.000000</td>
</tr>
<tr>
<td>1</td>
<td>.700000</td>
<td>3.800000</td>
<td>4.000000</td>
<td>-152.0000</td>
</tr>
<tr>
<td>2</td>
<td>.490000</td>
<td>4.340000</td>
<td>4.000000</td>
<td>584.8000</td>
</tr>
<tr>
<td>3</td>
<td>.463000</td>
<td>4.934000</td>
<td>-93.20000</td>
<td>-2182.160</td>
</tr>
<tr>
<td>4</td>
<td>.390100</td>
<td>4.205000</td>
<td>405.7600</td>
<td>8612.692</td>
</tr>
<tr>
<td>5</td>
<td>.517270</td>
<td>6.63014</td>
<td>-1384.664</td>
<td>-36118.36</td>
</tr>
<tr>
<td>6</td>
<td>.131629</td>
<td>-1.59249</td>
<td>5869.178</td>
<td>162372.5</td>
</tr>
<tr>
<td>7</td>
<td>1.43778</td>
<td>30.2795</td>
<td>-26669.53</td>
<td>-782851.6</td>
</tr>
<tr>
<td>8</td>
<td>-3.58970</td>
<td>-106.449</td>
<td>132819.5</td>
<td>404416.2</td>
</tr>
<tr>
<td>9</td>
<td>17.8758</td>
<td>535.079</td>
<td>-713389.2</td>
<td>-22335606</td>
</tr>
</tbody>
</table>

Mathematically, this awful behaviour is our punishment for interchanging the sum and integral in Eq. (15), and series such as these are called \textit{asymptotic series}. Fortunately, since the terms not only increase but also oscillate we can still assign a meaning to such sums by the procedure called \textit{Borel summation}, which in essence interchanges sum and integral back again. To see how this works, take a function

\[ f(x) = \sum_{n \geq 0} (-x)^{n} \alpha_{n} \quad (x \geq 0), \]  

(73)

where \( \alpha_{n} \) grows like \( n! \) for large \( n \) (as in the case of the perturbative expansion). Now, \( f(x) \) is not a convergent series, but

\[ g(x) = \sum_{n \geq 0} (-x)^{n} \frac{\alpha_{n}}{n!} \]  

(74)

is convergent, at least for some \( x > 0 \). Integrating term by term we can then see that a sensible definition of the value of \( f(x) \) is

\[ f(x) \equiv \int_{0}^{\infty} dy \ e^{-y} g(xy). \]  

(75)

\(^{16}\)The radius of convergence of a power series in \( z \) around \( z = 0 \) is defined as the largest value of \( |z| \) such that the power series converges, \textit{i.e.} the subsequent terms in the series decrease fast enough.
As an example, take, $\alpha_n = n!$; in that case
\begin{equation}
  g(x) = \frac{1}{1 + x}, \quad f(x) = \int dy \frac{e^{-y}}{1 + xy},
\end{equation}
and the resulting Borel-summed expression for $f(x)$ is well-defined\(^{17}\) for nonnegative $x$, while its expansion in powers of $x$ is not a convergent series. Note that for $x$ real and negative, the integral is ill-defined, and the terms in the series for $f(x)$ do not oscillate any more.

Perhaps of more direct computational relevance, asymptotic series like (73) have the property that if we truncate them at some order, the error made is smaller than the first neglected term, in absolute value:
\begin{equation}
  f(m)(x) = \sum_{n=0}^{m-1} (-x)^n \alpha_n, \quad |f(m)(x) - f(x)| \leq |x^m \alpha_m|.
\end{equation}
Hence, by choosing $m$ such that $|x^m \alpha_m|$ is minimal, we may hope to still get a good numerical approximation to the real answer. In our example, with $\alpha_n = n!$, the best $m$ is seen to be $m \sim 1/x$, and the error made is of the order of $m! x^m \sim \exp(-1/x)$, which for small $x$ is very small indeed! As long as we do not plan to compute the $137^{th}$ order, QED would appear to be safe, but in QCD the potential hazards are closer at hand, around the $10^{th}$ order. In any case it is good to keep in mind that perturbation theory is a crutch with cracks.

3. ENTER THE UNIVERSE: MANY POINTS MAKE A SPACE

3.1. One-dimensional theories

We now start to expand our model of the universe. While still keeping to a particle without intrinsic properties other than its mass, we shall allow it to live on more than just a single point. Moreover, we shall adopt the usual practice of choosing our units such that both $\hbar$ and $c$, the speed of light, have numerical value 1.

3.1.1. Theories with more field variables, and space

We shall start with a simple extension of our zero-dimensional model, and take an infinite set of fields, labelled by integer labels: $\varphi_n$, with $-\infty < n < \infty$. Since we are interested in expectation values of products of fields, like $\langle \varphi_2 \varphi_0 \varphi_1 \varphi_5 \rangle$, we also need to introduce an infinite number of sources $J_n$, one $J$ for every $\varphi$, since we use the $J$’s to identify and count the powers of the various $\varphi$’s. In the spirit of perturbation theory, we shall for now forget interactions that contain three or more $\varphi$’s. If the action only contained terms of type $\varphi_n^2$, we would end up with just an infinite number of copies of the zero-dimensional theory, and no correlations between fields would exist. Therefore, we take the action to be
\begin{equation}
  S(\{\varphi\}) = \sum_n \left( \frac{1}{2} \mu \varphi_n^2 - \gamma \varphi_n \varphi_{n-1} \right),
\end{equation}
and the path integral, the multi-variable generating function of the expectation values of all possible products of all possible powers of the various $\varphi$’s, becomes
\begin{equation}
  Z(\{J\}) = N \int \left( \prod_k d\varphi_k \right) \exp \left( -S(\{\varphi\}) + \sum_n \varphi_n J_n \right).
\end{equation}

\(^{17}\)For the experts: it is equal to $e^{1/x} E_1(1/x)/x$, where $E_1$ is the exponential integral.
Now we have interactions of nearest-neighbour type, and correlations between fields further away can only go via the intermediate fields: we expect the correlator

\[ \langle \varphi_m \varphi_n \rangle \equiv \Pi(n, m) = \Pi(|n - m|) \]  

(80)
to be smaller when $|n - m|$ is larger, and this is just what we expect for fields at large distances. In fact, it is the ‘link’ $\gamma$ that tells us that we might envisage the fields as living on a set of points in some space, the distance between neighbouring points becoming ‘larger’ as $\gamma$ decreases. **The neighbour-interactions tell us the structure of space.** This may at first sight appear mystical, but we should realize that we obtain our notions of space and time from ideas about how ‘distances’ between space-time events are defined: these are typically based on measuring correlations (for instance by sending light signals etcetera) and so they depend directly on the neighbour-interactions. In this simple case, since $\gamma$ and $\mu$ do not depend on $n$, we can envisage the points to be equidistant, separated by some ‘distance’ $\Delta$:

and the whole model is invariant under translation by one distance $\Delta$. Of course, in the end we shall take $\Delta \to 0$, and recover normal translation invariance.

### 3.1.2. The one-dimensional propagator

Let us write down the Feynman rules for our model. Since there are more fields we need to label each line:

\[ n = \frac{1}{\mu} \quad , \quad n = \gamma (\delta_{m,n+1} + \delta_{m,n-1}) . \]

(81)

The propagator $\Pi(n)$ is defined as the 2-point Green’s function $\langle \varphi_k \varphi_{k+n} \rangle$, and is given by Feynman diagrams with precisely two external legs, one corresponding to $\varphi_k$, the other one to $\varphi_{k+n}$, say. Because of translation invariance we may take $k = 0$. The SD equation for the propagator reads

\[ 0_0 = 0 + \sum_{k=\pm 1} 0_k , \]

(82)
or

\[ \Pi(n) = \frac{1}{\mu} \delta_{n,0} + \frac{\gamma}{\mu} (\Pi(n + 1) + \Pi(n - 1)) \]  

(83)

We can solve this by Fourier transform. Defining

\[ \Sigma(z) = \sum_n \Pi(n) e^{-inz} \quad \leftrightarrow \quad \Pi(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dz e^{inz} \Sigma(z) \]  

(84)

the SD equation becomes purely algebraic:

\[ \Sigma(z) = \frac{1}{\mu} + \frac{\gamma}{\mu} \Sigma(z) \left( e^{iz} + e^{-iz} \right) = \frac{1}{\mu - 2\gamma \cos z} . \]

(85)

We now take the *continuum limit*, $\Delta \to 0$, and see if we can obtain a sensible limit for the propagator. We keep the ‘distance’ between fields fixed, so we use $x = n\Delta$ with $x$ fixed. We also introduce the momentum as $p = z/\Delta$ such that $nz = px$. Note that then $\cos z = \cos(p\Delta) \sim 1 - p^2 \Delta^2/2$ for fixed $p$. 

19
The only sensible and nontrivial continuum limit occurs when we choose
\[ \gamma \sim \frac{1}{\Delta}, \quad \mu \sim \frac{2}{\Delta} + m^2 \Delta, \]
(86)
for small \( \Delta \), with some number \( m^2 \). The correlator then becomes\(^{18}\)
\[ R(x) \equiv \Pi \left( \frac{x}{\Delta} \right) \sim \frac{1}{2\pi} \int dp \frac{e^{ipx}}{p^2 + m^2} = \frac{1}{2m} e^{-m|x|}. \]
(87)
Note that, whereas the fields \( \varphi_n \) are not independent due to the \( \gamma \) interaction, the momentum modes are in fact independent: the description of particles in terms of states with given momentum is simpler than its configuration-space description, and we shall use momentum-space Feynman rules from now on.

Finally, it should be remarked that nearest-neighbour interaction gives the standard propagator, but there are suitably perverse choices including next-to-nearest neighbour or higher interactions that result in propagators with different behaviour.

3.1.3. The scalar action in one dimension

Let us now consider the action in the continuum limit. We can collect the whole configuration of values \( \{ \varphi \} = (\cdots, \varphi_{-1}, \varphi_0, \varphi_1, \varphi_2, \cdots) \) into a function \( \varphi(x) \), where \( \varphi(x) = \varphi_x / \Delta \). The action depends on this function, and is called a functional of the field \( \varphi(x) \). Carefully collecting terms, and rewriting \( \sum_n \Delta \cdots \) as \( \int dx \cdots \), we see that the limit becomes (cf. Eq. (78))
\[ S(\{ \varphi \}) = \sum_n \left( \frac{1}{2} \mu - \gamma \right) \varphi_n^2 + \frac{1}{2} \gamma (\varphi_{n+1} - \varphi_n)^2 \]
\[ \sim \sum_n \left( \frac{m^2 \Delta}{2} \varphi(x)^2 + \frac{1}{2\Delta} (\varphi(x + \Delta) - \varphi(x))^2 \right) \rightarrow \]
\[ \rightarrow S[\varphi(x)] = \int dx \left( \frac{1}{2} m^2 \varphi(x)^2 + \frac{1}{2} \varphi'(x)^2 \right) \equiv \int dx \mathcal{L}, \]
(88)
where we have defined the Lagrangian density \( \mathcal{L} \). A sensible limit for the source terms can only be obtained if we put
\[ J_n \rightarrow \Delta J(n \Delta) \quad \text{so that} \quad \sum_n \varphi_n J_n \rightarrow \int dx \varphi(x) J(x), \]
(89)
and the path integral reads
\[ Z[J(x)] = N \int \mathcal{D} \varphi \exp \left( -\int dx \left( \mathcal{L} - \varphi(x) J(x) \right) \right). \]
(90)
Here, \( \mathcal{D} \varphi \) is the ‘infinitesimal path integration element’, the limiting case\(^{19}\) of the earlier (enumerable) \( \prod_n \varphi_n \). A word on the behaviour of \( \varphi(x) \) is in order. From our interpretation of the action as describing a combined probability density for the field values \( \varphi_n \), it follows that the quantity \( (\varphi_{n+1} - \varphi_n)^2 / \Delta \) must be finite if a particular configuration of \( \varphi \)'s is to contribute to the path integral. So, steps in \( x \) of length \( \Delta \) typically make \( \varphi(x) \) jump with steps of order \( \sqrt{\Delta} \), in other words: in the limit \( \Delta \rightarrow 0 \) this \( \varphi(x) \) must be everywhere continuous but nowhere differentiable. Pictorially, a typical \( \varphi(x) \) has a fractal structure, with

\(^{18}\)Some care has to be taken here, since the approximation for the cosine is not valid everywhere. Fortunately, in the resulting integral only values of \( p \) of order \( 1/x \) contribute appreciably, and there the approximation is justified.

\(^{19}\)This limit is actually rather more subtle than presented here, since the number of continuous functions of \( x \) is not enumerable. For our purposes, however, we may skip such fine points.
zigs and zags on ever smaller lengths scales, like the path of a particle undergoing Brownian motion\textsuperscript{20}. The term $\varphi'(x)^2$ in the action has, therefore, a rather symbolic meaning.

3.1.4. The classical field equation in one dimension

For a single field in zero dimensions, the classical field equation reads $S' (\phi) = J$, with solution $\phi_c (J)$. This does not mean that this solution is the only one contributing to the path integral, but rather that a small interval (whose size diminishes with $\hbar$) of all $\varphi$ values around $\varphi_c = \phi_c (J)$ gives the dominant contribution. In the discrete one-dimensional case, the analogue of the classical field equation is

$$\frac{\partial}{\partial \varphi_k} S (\{ \varphi \}) = J_k \quad \forall k .$$

For the action (78), this results in

$$\mu \varphi_k - \gamma (\varphi_{k-1} + \varphi_{k+1}) = J_k ,$$

so the ‘equation of motion’ reads

$$\Delta m^2 \varphi (x) + \frac{1}{\Delta} (2 \varphi (x) - \varphi (x - \Delta) - \varphi (x + \Delta)) = \Delta J (x)$$

$$\Rightarrow m^2 \varphi (x) - \frac{1}{\Delta} \left[ \frac{\varphi (x + \Delta) - \varphi (x)}{\Delta} \right] - \frac{\varphi (x) - \varphi (x - \Delta)}{\Delta} \right] = 0 .$$

The second term is seen to be the discrete variant of a derivative. Denoting this discrete derivative by $E_x$, we see that the action in the discrete case can be written as

$$S (\{ \varphi \}) = \sum_n \Delta \left( \frac{m^2}{2} \varphi^2 + \frac{1}{2} (E \varphi)^2 \right) ,$$

and the equation of motion as

$$\frac{\partial}{\partial \varphi} S (\{ \varphi \}) - E_x \left( \frac{\partial}{\partial E \varphi} S (\{ \varphi \}) \right) = 0 .$$

In the continuum limit, the equation of motion becomes the Euler-Lagrange equation for the scalar action:

$$m^2 \varphi (x) - \varphi'' (x) = J (x) .$$

The classical solution $\phi_c$ is continuously differentiable for smooth source functions $J (x)$ and hence by itself is not even in the path integral measure. Again it is the set of $\varphi (x)$ functions close to the classical one that give the dominant contributions.

Derivatives such as the above, where the action functional is varied with respect to a single value of $\varphi$, are called functional derivatives, and usually denoted by curly $\delta$’s. In the text-book jargon, we may write the Euler-Lagrange equation as follows:

$$\left( \frac{\delta}{\delta \varphi (x)} - \frac{d}{dx} \frac{\delta}{\delta \varphi' (x)} \right) \left[ S [\varphi (x)] - \int dx \varphi (x) J (x) \right] = 0 .$$

\textsuperscript{20}In fact, for $m = 0$ it is Brownian motion in one dimension, where $\varphi (x)$ is the position at time $x$. 

21
3.2. More dimensions
Rather than bothering with the Feynman rules for the one-dimensional theory, we now move on to more dimensions. We shall always aim for the simplest case, and therefore stick to nearest-neighbour interactions. We shall start with a discretized action, where the field values are labelled by a set of several integers rather than a single one. Note that a labelling with several integers can always be re-encoded in terms of a single integer label: however, simple nearest-neighbour interactions in four dimensions would look very complicated in the one-dimensional encoding, leading to non-local interactions. Our notion of space – yes, the ‘everyday kind’ of space – is, after all, only a mental construction, developed over the course of our development up from foetuses to adults, that allows us to comprehend, and react to, the world in an efficient manner. **Spacetime is defined in such a way interactions look simple.**

3.2.1. Euclidean multidimensional theory
Let us consider the discrete version of a $D$-dimensional theory. Vector labels $\vec{n} = (n^1, n^2, n^3, \ldots, n^D)$ have been chosen for the fields. Also we define $\vec{e}_k$ to be the vector with $k^{th}$ component equal to 1, and the other ones 0. The action (again without self-interactions) is then given by

$$S(\{\varphi\}) = \sum_{\vec{n}} \left( \frac{1}{2} \mu \varphi^2_{\vec{n}} - \gamma \sum_{k=1}^{D} \varphi_{\vec{n}} \varphi_{\vec{n} + \vec{e}_k} \right).$$

(98)

There is now a translation invariance in each of the $D$ directions. The SD equation can be solved in the same manner as above, and we find for the correlator:

$$\Pi(\vec{n}) = \frac{1}{(2\pi)^D} \int_{-\pi}^{\pi} d^D \vec{z} \frac{\exp(i\vec{n} \cdot \vec{z})}{\mu - 2\gamma \sum_{k=1}^{D} \cos(z_k)},$$

(99)

where we have used the $D$-dimensional generalization of $t$. The continuum limit is now obtained by defining

$$\vec{x} = \Delta \vec{n}, \quad \vec{p} = \Delta \vec{p}, \quad \gamma \sim \Delta^{2-D}, \quad \mu \sim 2D\gamma + m^2 \Delta^D.$$

(100)

Using some algebra and standard integrals, we find for the continuum correlator

$$R(\vec{x}) = \frac{1}{(2\pi)^D} \int d^D \vec{p} \exp(i\vec{x} \cdot \vec{p}) \frac{1}{p^2 + m^2} = \frac{1}{2\pi} K_{1-D/2}(m|\vec{x}|) \left( \frac{2\pi|x|}{m} \right)^{1-D/2},$$

(101)

where $K$ is the modified Bessel function of the second kind. Note that we have obtained not only translational but also rotational invariance: the rectangular structure of the underlying discrete grid has become invisible in the continuum limit\(^{21}\). The continuum Lagrangian density is now just the more-dimensional generalization of Eq. (88):

$$\mathcal{L} = \frac{1}{2} m^2 \varphi(\vec{x})^2 + \frac{1}{2} \left( \nabla \varphi(\vec{x}) \right)^2.$$

(102)

In the same manner, the continuum source function is defined by $J_{\vec{n}} = \Delta^D J(\vec{x})$, and the classical field equation is, in the continuum limit:

$$m^2 \varphi(\vec{x}) + \nabla^2 \varphi(\vec{x}) = J(\vec{x}).$$

(103)

\(^{21}\)Again, the rotational invariance can be corrupted by choosing a more perverse action in the discrete case: fortunately, it is the simplest action that leads to the best continuum limit.
Finally, let us reintroduce the $\varphi^4$ self-interaction again. Cavalierly ignoring the fact that the propagator was derived from a SD equation without self-interaction vertices\footnote{In fact, the SD equation would be very different if we include the self-interactions, and so would the continuum limit. This is precisely the sin for which the loop divergences are the punishment.}, we simply add to the discrete action a term $\Delta^D \lambda \varphi_n^4$ for every $n$. The action including sources then becomes

$$S[\varphi(x)] = \int d^D x \left( \frac{1}{2} m^2 \varphi(\vec{x})^2 + \frac{1}{2} \left( \nabla \varphi(\vec{x}) \right)^2 + \lambda \varphi(\vec{x})^4 - J(x)\varphi(x) \right) ,$$  \hspace{1cm} (104)$$

Note that the $\lambda$ vertex also ‘occurs’ at all space points.

The Feynman rules for such a theory can be obtained in precisely the same manner as before, as the more-dimensional generalizations of the one-dimensional case. They are simplest in the momentum representation. For $D = 4$, we have the following Feynman rule for the propagator:

$$\frac{1}{p^2 + m^2} ,$$  \hspace{1cm} (105)$$

where we recognize the propagator of the one-dimensional theory in the momentum representation (with the replacement $p^2 \rightarrow \vec{p}^2$). For the single vertex in this theory, we have

$$-\lambda(2\pi)^4 \delta^4(p_1 + p_2 + p_3 + p_4) ,$$  \hspace{1cm} (106)$$

where the $(2\pi)^4 \delta^4()$ comes from the fact that we have to integrate this four-vertex over the whole space. Finally, the source is, in the same way, given by

$$-\lambda(2\pi)^4 \delta^4(p + q) ,$$  \hspace{1cm} (107)$$

In the vertices all momenta are counted incoming (or all outgoing). In addition, every momentum in a line has to be integrated over, with a factor $(2\pi)^{-4}$. Usually, we will consider sources that correspond to particles of fixed momentum being absorbed or produced (that is, sources $J(\vec{x})$ consisting of a single Fourier mode), and in that case the external momenta are effectively fixed.

3.2.2. Towards Minkowski space

The construction of our field theory looks almost realistic: the only thing not yet built in is the special rôle of time. Indeed, instead of space points labelled by $\vec{x} = (x^1, x^2, x^3, x^4)$, we know it is much better to describe things on the basis of space-time events with coordinates $x^\mu = (x^0, x^1, x^2, x^3)$, and a metric given by $g_{\mu\nu}$, with $g_{00} = 1, g_{11} = g_{22} = g_{33} = -1$, and the other components zero. We therefore make the following substitution:

$$x^4 = ix^0 .$$  \hspace{1cm} (108)$$

The exponent in the path integral, Eq. (104), then becomes

$$-S[\varphi(x)] \rightarrow iS[\varphi(x)] ,$$

$$S[\varphi(x)] = \int d^4 x \left( \mathcal{L} + J(x)\varphi(x) \right) ,$$

$$\mathcal{L} = \frac{1}{2} (\partial^\mu \varphi(x)) (\partial_\mu \varphi(x)) - \frac{1}{2} m^2 \varphi(x)^2 - \frac{1}{24} \lambda \varphi(x)^4 .$$  \hspace{1cm} (109)$$
The factor $1/24 = 1/4!$ is still there by convention. Note that we have taken a factor $-i$ out of the action integral. Again moving to the momentum representation:

$$\varphi(x) = \frac{1}{(2\pi)^4} \int d^4p \ e^{-ipx} \varphi(p) , \quad J(x) = \frac{1}{(2\pi)^4} \int d^4p \ e^{-ipx} J(p) ,$$  \hspace{1cm} (110)$$

where $\varphi(p) = \varphi(-p)$ and $J(p) = J(-p)$ since $\varphi(x)$ and $J(x)$ are real, the exponent can then be written as

$$iS = \sum_p \left( \frac{i}{2}(p^2 - m^2)\varphi(p)^2 + iJ(p)\varphi(p) \right. 
- \frac{i}{24} \lambda(2\pi)^4 \delta^4 (p_1 + p_2 + p_3 + p_4) \varphi(p_1)\varphi(p_2)\varphi(p_3)\varphi(p_4) \right) ,$$  \hspace{1cm} (111)$$

where the symbol $\sum_p$ means that every momentum has to be integrated over, with a factor $(2\pi)^{-4}$.

In the above, we have tacitly assumed that the $90^\circ$ Wick rotation implied by Eq. (108) is actually allowed: since the integrand in the path integral now is oscillatory rather than damped, this is not at all obvious. A signal of potential problems is the fact that the quadratic term in the action contains $p^2 - m^2 = (p^0)^2 - |\vec{p}|^2 - m^2$ which may become zero, and in fact does so precisely for on-shell particles! For the moment we shall blithely ignore this problem, and assume that all particles are off their mass shell; later on we will repair this.

### 3.3. Choosing the action

Until now we have mainly been studying $\varphi^4$ theory as a useful vehicle to develop the various techniques and concepts. Interactions such as $\varphi^3$ or interactions containing derivatives can easily be incorporated, the Feynman rules being read off in each case directly from the Lagrangian density: but which density ought we to take? A possible answer comes from the classical field equation, which as we have seen is the classical Euler-Lagrange equation for the action. In the $\varphi^4$ case, it reads

$$\left( \frac{\delta}{\delta \varphi(x)} - \frac{\partial}{\partial x_{\mu}} \frac{\delta}{\delta \partial^\mu \varphi(x)} \right) \mathcal{L} = 0 \quad \Rightarrow \quad \partial^\mu \partial_\mu \varphi(x) + m^2 \varphi(x) + \frac{\lambda}{6} \varphi(x)^3 = J(x) .$$  \hspace{1cm} (112)$$

If we consider the particles `on their own", that is, without self-interaction or sources, this becomes

$$\left( \partial^\mu \partial_\mu + m^2 \right) \varphi(x) = 0 ,$$  \hspace{1cm} (113)$$

which is the Klein-Gordon equation. This gives us the necessary hint: to see what the propagator of a particle type ought to be, look at its `classical' field equation. The particle we have been studying so far was assumed to have no internal structure: its only characteristic can be its momentum, and the Klein-Gordon equation just says that for a free particle this momentum must be on the mass shell. For the vertices we may choose anything, but the burden of proof is of course ours: apart from the obvious conformity with experiment, there are some general considerations, for instance the resulting connected Green’s functions had better be Lorentz invariant. There is, however, another fundamental requirement, which we shall come to in the next Section.
4. QUANTUM TERRORISM: OF SCATTERING, CUTTING, AMPUTATING AND EXPLODING

4.1. The scattering and probability

What we have done so far would be doomed to be no more than a mathematical recreation unless we attempt to make contact with physics. In quantum field theory, the connected Green’s functions are postulated to relate to the quantum-mechanical transition amplitudes, and in that way enter into predictions for cross sections and lifetimes. The precise relation will follow below: first we need to make some general remarks.

4.1.1. The S matrix and the T matrix

In particle physics experiments, we typically start by preparing an initial state at some time ‘far in the past’ where particles are ‘far away from each other’: so far away, in fact, that we assume the particles to be free of any interaction. At that moment, the initial state \(|\text{in}, t = -\infty\rangle\) is then made up (ideally!) as a combination of single-particle states corresponding to, say, the incoming \(e^+\) and \(e^-\) each with its own momentum. As time takes its course, the particles will approach and start interacting. The incoming state evolves in some complicated manner, and at a late time it will be \(|\text{in}, t = +\infty\rangle\). It is important to realize that this is still the same state; it is rather, the experimentalist observer at \(t = +\infty\) who has undergone a time translation with respect to the accelerator operator who prepared the state at \(t = -\infty\). Then we perform the observation, and the state is observed to consist of, say, quarks moving away from the interaction region (we cheerfully ignore hadronization here). The corresponding state is \(|\text{out}, t = +\infty\rangle\). The probability (amplitude) for this to happen is of course the overlap between the states: \(\langle \text{out}, t = +\infty | \text{in}, t = +\infty \rangle\). Under the assumption that the prepared state is, at \(t = -\infty\), essentially that of a combination of free particles without interaction, and the observed state at \(t = +\infty\) likewise, the amplitude may be written as the element of a matrix connecting these free-particle states. This matrix is called the S matrix, and its elements are what we want to compute.

4.1.2. Unitarity and the optical theorem

If the free-particle states form a complete set, an important consequence follows. Let us denote by \(j\) a state that looks like a free-particle state at \(t = -\infty\): \(j\) therefore labels a particle content as well as all kinds of momenta, spins, etc.; and let us similarly denote by \(k\) a state that looks like a free-particle state at \(t = +\infty\). The element of the S matrix can then be written as \(S_{kj}\). The completeness of the whole set of possible states \(j\), and of the whole set of possible states \(k\), then implies that \(S\) must be unitary:

\[
SS^\dagger = S^\dagger S = 1 .
\]  
(114)

This requirement of unitarity essentially means that the normalization of states is preserved, in other words: **unitarity = conservation of probability**. If there were no interactions whatsoever, \(j\) could of course only appear as \(j\) even at \(t = +\infty\), and we would have \(S_{kj} = \delta_{k,j}\). It therefore makes sense to write the S matrix as follows:

\[
S = 1 + iT ,
\]  
(115)

where the \(i\) is taken out of the \(T\) matrix by convention. The unitarity of \(S\) in Eq. (114) then implies for \(T\):

\[
i(T - T^\dagger) + TT^\dagger = 0 .
\]  
(116)
Let us now specialize to the case where \( j = k \), i.e. the final and initial state happen to be identical, ‘in spite of’ all the interactions that have taken place. Eq. (116) then reduces to

\[
\text{Im}(T_{jj}) = \frac{1}{2} \sum_n |T_{nj}|^2,
\]  

(117)

where on the right-hand side we must sum over all states \( n \): the (imaginary part of) the amplitude \( T_{jj} \) is related to the total probability for the initial state \( j \) to go into any final state. Equation (117) is called the optical theorem since it was first derived in the context of waves travelling through a medium.

Another consequence of unitarity is that \( S \) (and \( T \)) matrix elements cannot be arbitrarily large. Indeed, if all labels are discrete, we must have \( \sum_k |S_{kj}|^2 = 1 \). For continuous labellings (such as momenta) the condition is more involved, but the only thing we need here is the knowledge that unitarity will be endangered whenever we encounter matrix elements that show unlimited growth (for instance as a function of energy).

4.2. Towards predictions for experiments

The relation between Feynman diagrams and scattering processes is the postulate that the \( T \) matrix elements of the theory are related to the Green’s functions, that themselves can be computed using the Feynman diagrams of the theory. Note that we have ‘related to’, and not ‘equal to’: we shall discuss the precise nature of the relation below. We shall not give rigorous derivations here\(^23\): instead, we aim to illustrate the consistency of our approach.

4.2.1. Density of states

The first thing to be done is to give a more precise meaning to the notion ‘sum over states’ as used for instance in Eq. (117). What we use for the sum over states is of course to some extent arbitrary, since we have not discussed the normalization of the states. An important requirement, however, would seem to be that the sum over states must be Lorentz-invariant, otherwise it becomes difficult to arrive at Lorentz-invariant predictions for cross sections and the like. Realizing that we assume the final-state particles to be free and on their mass shell, and to have positive energy\(^24\), the phase-space integration element for a single particle with momentum \( p^\mu \) and mass \( m \) is best defined to be

\[
\frac{1}{(2\pi)^3} d^4 p \frac{\delta}{\delta (p^2 - m^2)} \theta(p^0) = \frac{1}{16\pi^3} d^3 \vec{p} \frac{1}{\sqrt{\vec{p}^2 + m^2}}, \quad (118)
\]

where the first alternative is explicitly Lorentz-invariant. The factor \((2\pi)^{-3}\) comes from the usual \((2\pi)^{-4}\) that comes with any momentum integral, and a factor \(2\pi\) that comes with any Dirac delta function – but again, this is to some extent arbitrary. For a total final state consisting of \( n \) particles we of course also have the requirement of overall conservation of energy and momentum, and therefore we choose as the actual \( n \)-particle phase space integration element:

\[
dV(P; p_1, \ldots, p_n) \equiv (2\pi)^4 \delta^4 \left( P - \sum_{j=1}^n p_j \right) \prod_{k=1}^n \left[ \frac{1}{(2\pi)^3} d^4 p_k \frac{\delta}{\delta (p_k^2 - m^2)} \theta(p_k^0) \right], \quad (119)
\]

where \( P^\mu \) is the total four-momentum. This is defined to be the ‘sum over states’ for particles without characteristics other than momentum. If additional quantum numbers such as spin or colour etc. are involved, we include discrete sums over these variables as well: usually we shall leave them to be understood.

\(^23\) After all, no-one can prove that the world is described by quantum mechanics. \textit{Rigor mathesis perfectus mortis rigorem inducit.}

\(^24\) The on-shell condition is essentially the \textit{definition} of what it means to be a free particle with no energy other than that due to its motion (i.e. no potential energy); the positivity of the energy is an empirical fact.
4.2.2. Cross sections and widths

In making experimental predictions we have to obtain, in one way or the other, the quantum-mechanical transition amplitude $\mathcal{M}$ for some process. Presently, we shall see how $\mathcal{M}$ is related to the Feynman diagrams and the $S$ matrix element. Having obtained $\mathcal{M}$, we have to turn it into an experimental prediction. The two commonly studied quantities are the cross section for $2 \rightarrow n$ processes, and the decay width (inverse lifetime) for $1 \rightarrow n$ processes. At this point, we simply give the prescriptions for them. By $\langle |\mathcal{M}|^2 \rangle$ we denote the absolute value squared of $\mathcal{M}$, summed over unobserved discrete quantum numbers of the final state such as spins or colours, and averaged over similar quantities for the initial state$^{25}$.

For the partial decay width $d\Gamma$ of an unstable incoming particle with mass $M$ and momentum $P$ we have

$$d\Gamma = \frac{1}{2M} \langle |\mathcal{M}|^2 \rangle \, dV(P; p_1, \ldots, p_n) \, S_{BF},$$

and for the partial cross section $d\sigma$ for two incoming particles of masses $m_a$ and $m_b$ and momenta $p_a^\mu$ and $p_b^\mu$, respectively, we have

$$d\sigma = \frac{1}{2F} \langle |\mathcal{M}|^2 \rangle \, dV(P; p_1, \ldots, p_n) \, S_{BF},$$

$$F = \left( (s - m_a^2 - m_b^2)^2 - 4 m_a^2 m_b^2 \right)^{1/2},$$

$$s = P^2, \quad P^\mu = p_a^\mu + p_b^\mu.$$  \hspace{1cm} \hspace{1cm} (121)

Here, $S_{BF}$ is the Bose-Fermi symmetry factor: if the final state contains a group of precisely $p$ indistinguishable particles in the final state, we have to include a factor $1/p!$ for each such group, since if the particles are indeed indistinguishable, any permutation of them leads to overcounting in the cross section$^{26}$.

The ‘flux factor’ $1/2M$ is also related to our normalization of states, and we shall have to prove its consistency. That the relative ‘flux factors’ $1/2M$ and $1/2F$ are consistent with one another can be seen from the fact that, in the rest frame of particle $b$, we have

$$2F = 4m_b|\vec{p}_a| = (2m_b) \left( 2m_a \cdot \frac{E_a}{m_a} \right) \left( \frac{|\vec{p}_a|}{E_a} \right),$$

where we recognize the normalization of the two states (with the wave of $a$ having undergone the correct Lorentz contraction in its direction of motion, and hence a change in its normalization with respect to that for a particle at rest), and the relativistic velocity of $a$ in the rest frame of $b$, inherent in the definition of cross section as the rate per unit flux.

Finally, dimensionality. Taking into account$^{27}$ that the Dirac delta function $\delta(A)$ has the inverse dimension of its argument $A$, the $n$-particle phase space factor is easily seen to have dimension $(\text{GeV})^{3n-4}$. If an inverse lifetime is to have dimension (GeV)$^4$, then $\mathcal{M}(1 \rightarrow n)$ must have dimension (GeV)$^{3-n}$. We conclude that an amplitude with a total of $k$ external particles must have dimension (GeV)$^{4-k}$. It follows immediately that cross sections must have dimension (GeV)$^{-2}$, in accordance with their interpretation as an area. Note that we deal here with the units in which the amplitudes are expressed, and not with their actual behaviour as a function of the total energy: indeed, this is what leads into trouble later on.

$^{25}$This means, for instance, a factor $1/2$ for each unpolarized incoming spin-$1/2$ particle, $1/3$ for a massive spin-$1$ particle, $1/2$ for an incoming massless spin-$1$ particle, $1/3$ for an incoming quark, and $1/8$ for an incoming gluon.

$^{26}$On the other hand, the number of Feynman diagrams in $\mathcal{M}$ will be bigger by a factor $p!$. Some authors include a factor $\sqrt{S_{BF}}$ in the definition of $\mathcal{M}$, but this spoils some of the symmetry in case one of the incoming particles is indistinguishable from one of the outgoing ones.

$^{27}$as can be easily seen from the definition $\int dA \delta(A) = 1$. 

27
4.2.3. Connected and disconnected diagrams

We are now in a position to take Feynman diagrams seriously as saying something about physics going on. At this point, the connectedness of the diagrams starts to play a significant rôle. Consider the diagrams associated with the decay of a single particle: they contain one incoming, and any number of outgoing particles. Some of these diagrams may be connected, but in principle we can also write down disconnected diagrams, as indicated below, where as usual hatched blobs denote connected diagrams, and here a speckled blob denotes any combination of connected and disconnected diagrams:

If the outgoing particles all have positive energy, we see that the connected diagrams without the incoming line attached vanish because of energy conservation. Therefore, in this case only the connected Feynman diagrams can give a nonzero result. Similarly, for the two-particle cross section, the associated Feynman diagrams can be schematically given as

Again, blobs with no incoming lines vanish because of energy conservation.

The case with two separate connected contributions, each with one incoming line, is more tricky. It can contribute, but only if the two incoming particles are capable of decaying independently in the particular indicated final states. In most cases this will not be possible (for instance in $e^+e^-$ collisions), and then the conclusion is that only the connected Feynman diagrams contribute. This is, in fact, the reason for the emphasis on connected diagrams in our zero-dimensional model. It must be kept in mind, however, that this is not really a theorem. Indeed there are cases where the disconnected diagrams do enter. For example, we may have collisions between unstable particles such as $\mu^+\mu^-$, or between, say, three or more particles. Another instance may be processes occurring at high temperature (such as the very early universe), where the surrounding ‘heat bath’ can actually supply energy ‘out of the vacuum’ such that particles may be created without explicit incoming particles. Fortunately, in typical collider situations such exotic processes can usually be neglected.

Having argued that only the fully connected diagrams give a physical contribution, we make one more step. Each such diagram contains contains an overall factor $(2\pi)^4\delta^4(P - \sum p)$, where $\sum p$ stands for the sum of all outgoing momenta, and $P$ for the total incoming momentum. The square of a Dirac $\delta$ function is not well-defined, so as usual, we enclose our system in a large space-time box, thereby turning the Dirac $\delta$ function into a simpler Kronecker $\delta$, which can be squared. Letting the box become infinitely large again at the end, we recover precisely the overall momentum conservation that we already included in our sum over states, times volume factors that cancel against the wave function normalizations. Part of our recipe, then, is that we should sum all the relevant Feynman diagrams, but leave out the overall momentum-conservation factor.

4.3. Poles and residues

We are not yet able to give a precise recipe for the matrix element $\mathcal{M}$, since the propagator $i/(p^2 - m^2)$ is still singular precisely for stable, on-shell particles. We are now ready to face this issue: its resolution also tells us the correct prescription for $\mathcal{M}$.
4.3.1. The propagator for an unstable particle

Let us consider the following process. We start with the vacuum, with no particles at \( t = -\infty \). Then, a source emits an unstable particle with momentum \( p^\mu \) and mass \( m \), which propagates through spacetime, while undergoing interactions which may destroy the particle by letting it decay, but do not necessarily do so. Let us assume that the particle is not destroyed, but rather absorbed by another source. Then, at \( t = +\infty \) we are left with the vacuum again. Let us denote the particle’s actual propagator by \( iR(p) \). At this point we do not know the precise form of \( R(p) \) except that it should look like \( 1/(p^2 - m^2) \) if \( p^\mu \) is sufficiently far off its mass shell. Putting in the source vertices as well we have

\[
iT_{kk} = \begin{array}{c} \hline \hline \end{array} \begin{array}{c} \hline \hline \end{array} \begin{array}{c} \hline \hline \end{array} = -iJ^2 R(p) ,
\]

where the blob denotes all possible connected diagrams. The label \( k \) stands here for the vacuum: both before the source emits the particle, and after it absorbs the particle again, the states are the same, and we are indeed looking at a diagonal element of the \( T \) matrix. Now we apply the optical theorem: the imaginary part of \( T_{kk} \) is related to \( \sum_n |T_{nk}|^2 \), where \( n \) denotes all possible final states. Pictorially, this sum reads

\[
\frac{1}{2} \sum_n |T_{nl}|^2 = \frac{1}{2} \sum_n \left| \begin{array}{c} \hline \hline \end{array} \begin{array}{c} \hline \hline \end{array} \begin{array}{c} \hline \hline \end{array} \right|^2 ,
\]

where the half-open blob denotes any decay process. The sum over \( n \) is nothing but our ‘sum over states’, summed over all possible particles configurations that could possibly result from the decay of our unstable particle. In other words, we must have a relation of the form

\[
\frac{1}{2} \sum_n |T_{nl}|^2 = J^2 |R(p)|^2 k m^{\Gamma_{tot}},
\]

if \( p^\mu \) is close to its mass shell, where \( \Gamma_{tot} \) is the total decay width of the particle. This relation simply follows from the definition of \( \Gamma_{tot} \). At this point we do not know yet what \( k \) is, since that is precisely one of the factors in the transition from diagrams to \( M \) that we want to determine; however, we do know that it must be a positive dimensionless number. At any rate, we have the following condition for \( R \) from the optical theorem:

\[
-\text{Im}R(p) = k m^{\Gamma_{tot}} |R(p)|^2 .
\]

Together with the desired form for \( R \) far from the mass shell, the simplest solution is to take

\[
R(p) = \frac{1}{p^2 - m^2 + ikm^{\Gamma_{tot}}} .
\]

A more precise analysis would take into account that the invariant mass at which the decay process is evaluated is not \( m \) but rather \( \sqrt{p^2} \), and the imaginary part is accordingly \( p^2 \) dependent: the above form holds for narrow resonances\(^{28} \).

4.3.2. The \( \text{i} \varepsilon \) prescription

What, now, for stable particles? The simplest idea is to consider stable particles as unstable particles with an extremely long lifetime. In that case, we just let \( \Gamma_{tot} \) decrease down to almost, but not quite, zero. The precise value of \( \Gamma_{tot} \) is then not relevant anymore, only the fact that \( km^{\Gamma_{tot}} \) should be infinitesimally small but positive, and hence our Feynman rule for stable particles reads

\[
\begin{array}{c} \hline \hline \end{array} p \rightarrow \frac{i}{p^2 - m^2 + i\epsilon} , \quad \epsilon \rightarrow 0^+ .
\]

\(^{28}\)The \( p^2 \) dependence of the width does give measurable effects in, for instance, \( Z^0 \) production at LEP.
This is the famous $i\epsilon$ prescription. Note that the sign is very important: on it hinges the fact that unstable particles disappear, or in other words the direction of time: time is going forward in the direction in which unstable particles decay. Incidentally, the $i\epsilon$ prescription is also sufficient to allow us to do the Wick rotation in the path integral: it could have been put in from the start, by adding a small term $i\epsilon\sigma^2/2$ to the action, thus ensuring that the rotated path integrand vanishes at infinity.

4.3.3. Almost-stable particles and amputation

We have seen that a good description of stable particles is to consider them as almost-stable unstable ones. In the $S$ matrix, however, unstable particles never make it to $t = +\infty$, and we have to take a closer look at what is going on precisely. Let us assume that in some process an unstable particle is produced off-shell, with momentum $p$. Having propagated for a while it must decay, but if the particle is not very unstable this may be quite a distance away. Diagrammatically, we have

\[
\begin{array}{c}
A \quad \{k\} \\
\quad \quad p \\
\quad \quad B \quad \{q\}
\end{array} + \text{other diagrams},
\]

where $A$ denotes the connected graphs in the ‘production’ process, and $B$ those in the ‘decay’ process. The $\{q\}$ and $\{k\}$ stand for the other particle momenta involved. As indicated, there may be other diagrams leading to the same final state, but without involving this unstable particle. Let us denote by $A$ the amplitude (a number) corresponding to $A$, and by $B$ that corresponding to $B$. The cross section, according to the postulated rules, is given by

\[
d\sigma = \frac{1}{2F}|A|^2 \frac{1}{(p^2 - m^2)^2 + k^2m^2\Gamma_{\text{tot}}^2} |B|^2 dV(P; \{q\} + \{k\}) ,
\]

plus other terms that do not contain the $p$ propagator squared. We now multiply by unity in the form $(2\pi)^{-4}d^4p(2\pi)^4\delta^4(p - \{q\})$, and we get

\[
d\sigma = \frac{1}{2F}|A|^2 dV(P - p; \{k\}) \frac{1}{(2\pi)^4} \frac{d^4p}{(p^2 - m^2)^2 + k^2m^2\Gamma_{\text{tot}}^2} |B|^2 dV(p; \{q\}) ,
\]

plus the other terms, which peak less strongly.

Let us now take the limit of almost-stability. The propagator then becomes very sharply peaked, and we may write

\[
\frac{1}{(p^2 - m^2)^2 + k^2m^2\Gamma_{\text{tot}}^2} \sim \frac{\pi}{km\Gamma_{\text{tot}}} \delta\left(p^2 - m^2\right).
\]

The normalization can be simply checked by integration over $p^2$. Putting this in, the expression for the cross section becomes

\[
d\sigma \sim \frac{1}{2F}|A|^2 dV(P; p + \{k\}) \frac{1}{\Gamma_{\text{tot}}} \frac{1}{2mk} |B|^2 dV(p; \{q\}) ,
\]

where we have neglected the non-resonant parts since they cannot contribute in the infinitesimally small part of phase space where $p^2 = m^2$.

The last factor in the above expression is nothing but what we are supposed to be using to compute the partial decay width $p \rightarrow \{q\}$, including the as yet unknown factor $k$. The result is exactly what we ought to expect on the basis of probability theory, namely that the probability to produce an almost-stable particle, followed by its decay into a given final state is the product of the probability of producing it, times the branching ratio for the particular decay. Since stable particles are supposed to be decaying...
infinitely far away, and we cannot see them decaying, we have to sum over all their possible decay modes.

The resulting rule is attractively simple: **external legs corresponding to scalar particles must be replaced by the factor 1.** This ‘cutting off’ of external legs goes by the grisly name of **amputation.**

Before we finish this Section, the factor $k$ has to be cleared up. Since we have decided, on the basis of the probability interpretation of the cross section, that external scalar legs should be simply amputated, the same should of course hold for the incoming legs as well, and hence we should have $k = 1$.

### 4.4. Cutting rules

Having established how external particles have to be treated, let us return to the optical theorem. Denoting the sum of all Feynman diagrams that lead from an initial state $j$ to a final state $k$ by $D_{kj}$, the fact that the $S$ matrix element is $S_{kj} = \delta_{kj} + D_{kj}$ leads to the following expression of unitarity:

$$D_{kj} + D_{jk}^* + \sum_n D_{nj} D_{nk}^* = 0.$$  \hfill (134)

If perturbation theory is any good, Eq. (134) should hold term by term in an expansion in powers of all the coupling constants in the theory. Each such term may consist of many Feynman diagrams, and we shall argue that Eq. (134) in fact holds for each individual diagram.

To see how this works, consider the following diagram:

![Diagram](image)

This diagram comes from a theory with a single field and interaction potential

$$V(\varphi) = \frac{1}{6} \lambda_3 \varphi^3 + \frac{1}{24} \lambda_4 \varphi^4,$$  \hfill (135)

and is of order $\lambda_3^3 \lambda_4^2$. There are a good many other diagrams of the same order, and Eq. (134) must hold for their sum. Now assign a different label to each line in the diagram, for instance

![Diagram](image)

Let every line labelled $k$ now correspond to a field $\varphi_k$. Consider now a theory for 10 different fields $\varphi_1, \ldots, \varphi_{10}$, with interaction potential

$$V(\varphi_1, \ldots, \varphi_{10}) = g_1 \varphi_1 \varphi_2 \varphi_3 + g_2 \varphi_2 \varphi_4 \varphi_5 \varphi_6 + g_3 \varphi_3 \varphi_4 \varphi_5 \varphi_7 + g_4 \varphi_6 \varphi_8 \varphi_{10} + g_5 \varphi_7 \varphi_9 \varphi_{10}.$$  \hfill (136)

It is easy to check\(^{29}\) that for this theory, the contribution to $1 \rightarrow 8 + 9$ of order $g_1 g_2 g_3 g_4 g_5$ contains precisely this diagram, and only this one. Assuming this theory to be unitary as well, we see that Eq. (134) must hold for this diagram individually. What about the summed-over state $n$ in this case? It is easy to check that it can only be $2+3$, or $2+4+5+7$, or $3+4+5+6$, or $6+7$: every other choice would give the wrong order in the coupling constants. Diagrammatically, we can denote this by ‘cutting’ the diagram in half.

\(^{29}\)The secret is that all external lines occur in precisely one interaction term, and the internal lines in precisely two.
along some curve. The four possibilities in this case are given here:

A similar argument can be constructed for every Feynman diagram. The sum over $n$ is then a sum over all possible cuts that separate $j$ from $k$. The internal lines that are cut through are then ‘put on the mass shell’, and we have to sum/integrate over all their quantum numbers and momenta. Eq. (134), applied to individual diagrams, are called the cutting rules. Two remarks are in order about our treatment. In the first place, one might have worried about the symmetry factors, since diagrams with non-identical fields have symmetry factor 1, while the same diagram has another factor (in this case, 1/2) if the fields are identical. In fact there is no inconsistency, precisely owing to our prescription for the Bose-Fermi factor $S_{BF}$ in the sum over states. Secondly, the above is not really a rigorous derivation, since we had to assume the unitarity of the 10-field theory as well: in fact one usually first proves the cutting rules from the Feynman rules, and then proves the unitarity of $S$ from the cutting rules. Here, we merely aimed at showing the mutual consistency of unitarity and cutting rules.

4.5. Appendix: physical divergencies?

We have seen how a careful treatment of stable particles as almost-stable unstable ones led us to handle the apparent singularity in their propagators for external lines. But what about internal lines? That there may actually be singularities inside the physical phase space can be seen from the following simple model.

Take a model with three scalar particle species: $a$, $b$ and $c$. Let us assume that the following decays are physically possible: $a \rightarrow bb$ and $c \rightarrow ab$. This means, of course, that some (effective-action) vertices of the type $\varphi_a \varphi_b^2$ and $\varphi_a \varphi_b \varphi_c$ exist, and also that $m_c > m_a > m_b$; let us take $m_b = 0$ for simplicity.

Now consider the process

$$a(p_1) + a(p_2) \rightarrow b(p_3) + c(p_4),$$

where we have indicated the particles’ momenta. One of the diagrams contributing to this process is

where we have indicated with arrows the direction in which positive energy has to flow in this diagram. It is matter of simple two-body kinematics to check that the momentum $q^\mu$ of the internal $b$ line can become lightlike, $q^2 = 0$, if the total invariant mass squared $s$ is larger than $m_c^4/(m_c^2 - m_a^2)$: for each such an energy there is precisely one scattering angle (in the CM frame) for which this happens. For generic couplings, the amplitude $M$ and the cross section will therefore display a real, non-integrable singularity inside the physical phase space!

The way out of this conundrum is to realize that if the kinematics are such that the internal $b$ can be on its mass shell, we have here the decay of the $a$ particle inside a scattering process. If the non-stability of the $a$ is relevant inside the scattering, then we should of course also keep track of it for the external lines: an unstable incoming $a$ cannot come from infinitely far away, and its momentum is not fixed at

---

30There is, of course, also the diagram with $a(p_1)$ and $a(p_2)$ interchanged: that diagram also has the divergence, but at another scattering angle. We can therefore consider the diagrams separately as long as we only study their singularity structure.
\[ p_1^2 = m_1^2, \text{ but rather has a distribution around this value, with a width given by the } a \text{ lifetime. Instead of} \]

a pure momentum state for the incoming \( a(p_1) \), we ought therefore to use a superposition of states with different invariant masses. This ‘smearing’ turns out to be precisely sufficient to dampen the singularity in \( M \), and the cross section turns out to factorize, in a similar way as above, into a part describing the production of an on-shell \( b \) with momentum \( q^\mu \) in the decay of \( a(p_1) \), followed by the collision of \( b(q) \) and \( a(p_2) \) into \( c(p_4) \). The message, again, is: be wary of particles that claim to be stable.

5. INNER LIFE: PARTICLES WITH SPIN

So far, we have considered only scalar particles. These are, by definition, characterized by their momentum only. In the real life of the Standard Model, almost all particles have more structure than that. Charged particles are described by complex rather than real fields, opening up the possibility of antiparticles. The charged \( \pi \) is an example, although not an elementary one. Particles with additional internal spin structure are represented by fields with several components that get mixed in well-defined ways under Lorentz transformations, and coloured particles have in addition colour components that are mixed under complex rotations in colour space.

5.1. Dirac fermions

One of the most important types of spinning fields is that of spin-1/2 particles, of which the electron is the simplest type. We shall not, in these notes, go through the whole construction of Dirac spinors and their interpretation, but jump nimbly to the Feynman rules that describe them.

5.1.1. The Dirac equation and spinors

In his search for a relativistic wave equation that, in contrast to the Klein-Gordon equation (first put forward by Schrödinger), would be of first rather than second order in time, Dirac had to construct a set of matrices \( \gamma^\mu (\mu = 0, 1, 2, 3) \) that obey an anticommutation relation:

\[
\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu} \mathbf{1},
\]

where \( \mathbf{1} \) denotes the unit matrix. The matrix \( \gamma^0 \) must be hermitian, and the other three anti-hermitian. The simplest such set consists of \( 4 \times 4 \) matrices. Together with the unit matrix, these form the basis of the Dirac algebra. Each element of this algebra can be written as a sum of products of various \( \gamma \) matrices. Especially the total product is important: we use

\[
\gamma^5 \equiv i\gamma^0 \gamma^1 \gamma^2 \gamma^3 \Rightarrow (\gamma^5)^2 = \mathbf{1}, \quad \gamma^5 \gamma^\mu = -\gamma^\mu \gamma^5.
\]

The Dirac wave equation for a free electron reads

\[
(i\nabla + m)\psi(x) \equiv (\gamma^\mu \partial_\mu + m)\psi(x) = 0,
\]

where we introduced the notation \( \partial = \gamma^\mu a_\mu \) for any vector \( a \), so that \( \partial \partial = a^2 \). The wave function \( \psi \) has to have 4 components, two of which are interpreted as the two spin states of the electron \( e^- \), and the remaining two are related to the two spin states of the positron \( e^+ \). Multiplying Eq. (139) on the left with \( (i\nabla - m) \) we recover the mass-shell condition \( p^2 = m^2 \) for free particles.

The full consequences of the Dirac equation, such as the proof that indeed it describes a spin-1/2 particle with an approximate gyromagnetic ratio \( g = 2 \), are not important here. What we do need are the algebraic conditions for the Fourier modes corresponding to a fixed, on-shell momentum \( p^\mu \) and a spin pointing in a given direction. The spin direction is described by the spin vector \( s^\mu \), and we require

\[
p^2 = m^2, \quad s \cdot p = 0, \quad s^2 = -1,
\]

33
so that in the rest frame of $p^\mu$, $s^\mu$ is just a direction in 3-space. Having these, the Dirac spinors are defined to be $u(p, s)$ for a particle and $v(p, s)$ for an antiparticle, and these are defined by their projection operators:

$$
u(p, s) = \frac{1}{2}(1 + \gamma^5 s)(\not{p} + m),$$

$$v(p, s) = \frac{1}{2}(1 - \gamma^5 s)(\not{p} - m),$$

(141)

where the Dirac conjugate $\bar{u}$ is defined as $\bar{u} = u^\dagger \gamma^0$. The normalization chosen here is our convention: in many texts, $1/2m$ is preferred to $1/2$. The two spin states for the particle and the antiparticle are related by $s^\mu \leftrightarrow -s^\mu$. We also have the perhaps more familiar spin sums

$$\sum_s u(p, s)\bar{u}(p, s) = \not{p} + m, \quad \sum_s v(p, s)v(p, s) = \not{p} - m.$$  

(142)

One remark is in order here. For very energetic fermions, the spin vector may blow up. If both $s^\mu$ and $\not{p}$ point in the same direction $\vec{e}$ in some frame, Eq. (140) gives

$$p^\mu = (E, p\vec{e}) \quad s^\mu = \pm \left( \frac{p}{m} \frac{E}{m} \vec{e} \right) \sim \pm \frac{1}{m} p^\mu + O\left( \frac{m}{E} \right),$$

(143)

where the last term holds in the high-energy limit. That this problem is only apparent can be seen from the projection operator in that case:

$$u(p, s)\bar{u}(p, s) = \left(1 \pm \frac{1}{m} \gamma^5 \not{p} + O\left( \frac{m}{E} \right) \right)(\not{p} + m)$$

$$= \frac{1}{2}(1 \pm \gamma^5 \not{p} + O(m)).$$

(144)

The resulting projection operator in the limit $m \to 0$ is that for helicity states. The divergence of $s^\mu$ with energy disappears owing to the Dirac algebra.

5.1.2. Fermion propagators

In the same way that the propagator of a scalar particle is related to the Klein-Gordon equation, the spin-1/2 Dirac propagator is (the momentum representation of) the Green’s function for the free Dirac equation. We find the following Feynman rule:

$$\frac{i}{p^\mu - m^2 + i\epsilon} = \frac{i}{p^2 - m^2 + i\epsilon}. \quad \text{(145)}$$

Note that this propagator is not even in $p^\mu$: the direction in which the momentum flows is important, and hence we orient the propagator with an arrow and count the momentum along the arrow. Notice the occurrence of a nontrivial numerator: it tells us that the various spin states propagate together through space. This becomes even more clear when we write, somewhat symbolically,

$$\frac{i}{p^2 - m^2 + i\epsilon} = \frac{i}{p^2 - m^2 + i\epsilon} = \frac{i}{p^2 - m^2 + i\epsilon},$$

although this is strictly speaking only correct on the mass shell, to which we shall now move.

31Strictly speaking these are not projection operators since the nonzero eigenvalue is not 1 but $2m$: the important point is that they provide a division of the Dirac space into four separate pieces.
5.1.3. Amputation for fermions

The above Feynman rule holds for stable fermions (hence the $i\epsilon$); for unstable fermions, we again have to insert the $m\Gamma^{\text{tot}}$. We can now repeat the arguments that led us to the amputation formula for scalar particles, with the added complication that the propagator has a matrix structure. Since a good amplitude $\mathcal{M}$ should have no Dirac indices hanging around, a fermion line in a diagram ought to have the form $\tilde{u}(\text{matrices})u$ or similar. The ordering of the components of a Feynman diagram becomes important now, and we adopt the usual convention of matrix multiplication, which means that the fermion lines are written down against the direction of the arrow.

Let us now assume that a fermion rather than an antifermion is produced and decays far away, as depicted for a scalar particle in Eq. (129). The amputation argument is completely analogous, since it relies only on the denominator of the propagators, and we see that the denominator of the fermion propagator is again amputated away, but the numerator is left. The spinor part $u(p)$ must now be assigned to the decay amplitude $B$, while the conjugate spinor part $\bar{u}(p)$ must go with the production amplitude $A$. The sum over spins in Eq. (146) is now justified, and indeed consistent since the spin of the propagating fermion is by assumption not detected (otherwise it would not be an internal line). A similar argument holds for the case of an antifermion. The upshot is that we may choose the following Feynman rules for external, on-shell fermions:

\begin{align}
\begin{array}{c}
p \rightarrow \tilde{u}(p, s) \quad \text{outgoing fermion}, \\
p \rightarrow v(p, s) \quad \text{outgoing antifermion}, \\
p \rightarrow u(p, s) \quad \text{ingoing fermion}, \\
p \rightarrow \bar{v}(p, s) \quad \text{ingoing antifermion},
\end{array}
\end{align}

(147)

where the momentum is defined to be physical, i.e. with positive energy\footnote{This convention allows us to decide whether we have, say, an outgoing fermion or an incoming antifermion in our process.}. In our derivation of the antifermion amputation rules we have neglected the minus sign in front of the factor $\sum v\bar{v}$: not unreasonably, since in the derivation we dealt with the cross section, hence with $|\mathcal{M}|^2$ rather than $\mathcal{M}$ itself. We might have decided to take, say, $-\bar{v}$ for an incoming antifermion in order to include this sign, but we prefer to stick to the more usual assignments. But the minus sign rears its head in other places, as we shall now discuss.

5.1.4. The Fermi minus sign

Let us take a graph in which two fermion lines appear in opposite directions, and let us consider the cut through these two fermion lines: this is depicted below.

The cut may also go through other internal lines. The momenta are $p^\mu$ and $q^\mu$, both with positive energy in the direction going from blob A to blob B (as required by the cutting rules). The two fermions are not necessarily of the same type and mass. According to our Feynman rules, the diagram before cutting
reads
\[ \text{Tr} \left[ B \frac{\hat{p} + m_1}{p^2 - m_1^2 + i\epsilon} A \frac{-\hat{q} + m_2}{q^2 - m_2^2 + i\epsilon} \right]. \]

The trace over spinorial indices is necessary because both $A$ and $B$ must have the form of Dirac matrices (possibly in a complicated combination). The Feynman rules we have extracted for external fermions, however, tell us to take, for the cut diagram:

\[ \text{Tr} [B u(p)\bar{u}(p)A v(q)\bar{v}(q)] = \bar{u}(p)A v(q)\bar{v}(q)B u(p), \]

where we have implied summation over the spins in the spinors. We see that these results are inconsistent by an overall sign because $\sum \bar{v}(q)v(q) = \hat{q} - m_2$. To repair this (and, hence, repair unitarity) we have to include an extra minus sign in the Feynman rules. The most economical way to do this is by two extra prescriptions.

By the structure of the Dirac spinors, every diagram consist of zero or more fermion lines connecting an external ‘outgoing spinor’, $\bar{u}$ or $\bar{v}$, with an ‘ingoing spinor’, $u$ or $v$, and zero or more fermion lines going in closed loops. All these are connected and dressed up by bosonic pieces. If we adopt the convention that in every diagram contributing to a certain $\mathcal{M}$ the ingoing external fermions, say, are put in a fixed order, the ordering of the outgoing is unambiguously given for each diagram, and these orderings may differ between different diagrams. The minus sign prescriptions can now be given as follows:

- Every closed fermion loop gives a factor $-1$.
- Every interchange of two outgoing external fermion lines (in the above sense) gives a factor $-1$.

That the second prescription is also necessary to maintain unitarity can be seen from the following example, where the solid lines without arrows stand for any bosonic particle:

The first diagram must, by the first rule, have a relative minus sign with respect to the second one when the diagrams are uncut: the second rule is necessary to ensure the same relative minus sign when the diagrams are cut.

Finally, it is important to observe that the Fermi minus sign holds for all fermions, not only those that you would naively consider as identical particles, such as two electrons: the two tree-level diagrams for Moeller scattering, $e^-e^- \rightarrow e^-e^-$, have a relative minus sign. Indeed, electron and positron are easily distinguishable (and we have a relative minus sign in the two tree-level diagrams for Bhabha scattering, $e^+e^- \rightarrow e^+e^-$), but the minus sign rules also hold for exotic interchanges such as between the muon neutrino and the top quark (somewhat less exotically, the two tree-level diagrams for $e^+e^- \rightarrow \nu_e\bar{\nu}_e$ also have a relative minus sign). Really, it depends on the set of interaction vertices whether or not one fermion can replace another in a Feynman diagram. In QCD for instance, there are no vertices that allow us to interchange a single up quark and a single down quark, and their relative minus sign is invisible; but if we take the weak interactions into account, extra diagrams and interchanges becomes possible, and the sign does matter.

5.1.5. Aside: long-distance communication

There is a difference between scalar and non-scalar particles. As we have seen from the amputation arguments, when a scalar particle is produced and decays very much later, the production and decay are separated in spacetime to such an extent that the only information going between them is the momentum of the scalar particle, and for the rest the particle’s decay can be considered to be independent from
its production. Not so for spinning particles: the amputation for, say, an outgoing fermion gives a factor \( \bar{u}(p) \), which is coupled to a factor \( u(p) \) in its subsequent decay: and both spin states of this spinor propagate together from production to decay. Therefore, more information than just momentum is exchanged between production and decay, no matter how far they are separated. Moreover, this will hold also for two particles that are produced in one point and subsequently move very far away from each other before decaying.

### 5.2. Massive vector bosons

#### 5.2.1. Polarization vectors and the Proca equation

We now turn to yet another particle type with internal structure, corresponding to spin-1. The best-known such particle is of course the photon, but it is simpler to start with massive spin-1 particles such as the \( W \) and \( Z \). The reason is that for such particles we can define a rest frame. Apart from their momentum \( p^\mu \), such particles are described by their polarization vector \( \varepsilon^\mu \), which in the rest frame of the particle is nothing but the direction in which the ‘electric’ field points, hence we have

\[
p^2 = m^2, \quad p \cdot \varepsilon = 0, \quad \varepsilon \cdot \varepsilon^* = -1,
\]

where \( m \) is again the mass. Note the similarity between these and Eq. (140). The last equation is written such that it holds for real as well as complex polarization vectors. In the massive case, there are 3 linearly independent solutions, corresponding to the fact that a massive spin-1 system has 3 different values that the spin component in a given direction can take. Real polarization vectors, such as \( \varepsilon^\mu_1 \), correspond to linear polarization, and complex ones, such as \((\varepsilon^\mu_1 + i\varepsilon^\mu_2)/\sqrt{2}\), to elliptic or even (as in this case) circular polarization. The sum over spins in the rest frame must read

\[
3 \sum_{j=1}^{3} \varepsilon^\mu_j (\varepsilon_j^*)^\nu = \begin{cases} 1 & \text{if } \mu = \nu \neq 0 \\ 0 & \text{otherwise} \end{cases}
\]

and therefore the spin sum in any Lorentz frame reads

\[
3 \sum_{j=1}^{3} \varepsilon^\mu_j (\varepsilon_j^*)^\nu = -g^{\mu\nu} + \frac{p^\mu p^\nu}{m^2}.
\]

The conditions on \( \varepsilon \) can conveniently be combined in the so-called Proca equation:

\[
\partial^\mu \partial_\nu W^\mu - \partial^\nu \partial_\mu W^\mu + m^2 W^\mu = 0,
\]

where \( \varepsilon_\mu \) is the coefficient of the mode of the \( W \) field with momentum \( p \). If we contract Eq. (150) with \( \partial^\mu \) we obtain \( \partial^\mu W^\mu = 0 \), or \( p \cdot \varepsilon = 0 \), and inserting this condition again the Eq. (150) gives the Klein-Gordon equation which takes care of the mass-shell condition for free particles. Note that this only works if \( m \neq 0 \).

#### 5.2.2. Vector boson propagators and amputations

Just as in the Dirac case, the propagator is related to the Green’s kernel of the field equation. The Greens kernel \( R \) must satisfy the Proca equation with a Dirac delta function on the right-hand side:

\[
\partial^\nu \partial_\nu R^\mu_\alpha (x) - \partial^\nu \partial_\mu R^\nu_\alpha (x) + m^2 R^\mu_\alpha (x) = \delta(x) g_{\mu\alpha},
\]

\[37\]

---

\[37\] The expression is correct in the rest frame, and it is Lorentz-invariant.

\[34\] The orthogonality of \( p \) and \( \varepsilon \), \( \partial \cdot W = 0 \), follows from contracting with \( \partial^\mu \), and the mass-shell condition then follows by inserting this condition again.
In momentum representation, we have of course $R_{\mu\nu}(p)$, and the Proca equation reads

$$-p^2R_{\mu\nu}(p) + p_\mu p_\nu R(p) + m^2 R_{\mu\nu}(p) = g_{\mu\nu},$$

(152)

with solution

$$R_{\mu\nu}(p) = \frac{-g_{\mu\nu} + p_\mu p_\nu/m^2}{p^2 - m^2}.$$ 

(153)

Notice again that this will not work for $m = 0$. As usual, for unstable particles a width, and for stable particles an $i\epsilon$ has to be added for reasons of unitarity. The Feynman rule for a stable spin-1 particle propagator therefore is seen to be

$$\mu \quad \rightarrow \quad \frac{i}{p^2 - m^2 + i\epsilon} \left( -g^{\mu\nu} + \frac{p_\mu p_\nu}{m^2} \right) = \frac{i}{p^2 - m^2 + i\epsilon} \sum_{j=1}^{3} \varepsilon_{j}^{\mu} \varepsilon_{j}^{\nu}.$$ 

(154)

Again, we recognize the spin-sum in the ‘symbolic’ last expression. The Feynman rules for external stable spin-1 particles follow again from the amputation argument:

$\begin{align*}
p, \, \varepsilon & \quad \rightarrow \quad \varepsilon^\mu, \\
\circ \quad \rightarrow \quad p, \, \varepsilon & \quad \rightarrow \quad \varepsilon^* \mu,
\end{align*}$

(155)

where the arrow shows the flow of positive energy. The index on the polarization vector has to be contracted with a corresponding vector index in whatever vertex the external line is attached to, just as in the Dirac case vertices involving fermions must carry a Dirac index. Since the propagator is even in the momentum we do not need to orient the spin-1 lines, nor is there a minus sign to worry about.

### 5.3. Massless vector bosons

#### 5.3.1. Longitudinal and transverse polarization

Since the photon, at least, is massless or almost so\(^{35}\), we need to consider the $m \rightarrow 0$ limit of the spin-1 case discussed so far. Here, we run into trouble. Consider a massive spin-1 boson moving along the $z$ axis with momentum $p$ and energy $E = \sqrt{p^2 + m^2}$. By boosting from the rest frame, we see that the three possible polarization vectors can be chosen as follows:

$$\begin{align*}
p^\mu & = (E, 0, 0, p), \\
\varepsilon_1^\mu & = (0, 1, 0, 0), \\
\varepsilon_2^\mu & = (0, 0, 1, 0), \\
\varepsilon_3^\mu & = \left( \frac{p}{m}, 0, 0, \frac{E}{m} \right) = \frac{p^\mu}{m} + O \left( \frac{m}{E} \right).
\end{align*}$$ 

(156)

The vectors $\varepsilon_1$ and $\varepsilon_2$ correspond to the field being at right angles to the velocity, and hence of course unaffected by the boost. They (and real or complex linear combinations of them) are called the transverse polarizations. The vector $\varepsilon_3$, however, corresponds to the field being in the direction of motion, and hence by Lorentz contraction becoming more and more intense as the velocity increases: this is called the longitudinal polarization. As the velocity approaches that of light when $m \rightarrow 0$, it blows up. That this is a problem can be seen from the Feynman rule for external particles: the matrix element $M$ contains $\varepsilon$ as an overall factor, so $M$ will blow up when $\varepsilon$ does, and we are bound for unitarity violation, especially when $m = 0$. Nevertheless, the propagators and external lines are (more or less) unambiguously determined.

\(^{35}\)The best limit to date is about $10^{-25}$ GeV.
5.3.2. Unitarity and current conservation

A possible way out of our dilemma is to realize that a theory contains more than propagators and external lines: there are also interaction vertices. If a particle behaves awfully but has no interaction with the rest of the cosmos, it is just the same as if it doesn’t exist, and we might as well ignore it. Let us consider a process where a very light and fast spin-1 particle is emitted. If it is transversely polarized, we do not worry about unitarity, so let us take it longitudinally polarized (denoted by \( L \)):

\[
p \varepsilon_L = M = M^\mu (\varepsilon_L)_\mu = M^\mu \left( \frac{p^\mu}{m} + \mathcal{O} \left( \frac{m}{E} \right) \right).
\] (157)

Here, we have written \( M^\mu \) for the rest of the diagram (or set of diagrams). We now adopt the law that the vertices in the theory have to be such that in all cases such as this, we have

\[
M^\mu p_\mu = 0.
\] (158)

If we can arrange this, then the longitudinal spin-1 particle decouples when \( m \to 0 \): it has no interaction with anything else in the world, and is effectively absent from the theory. In fact, the above requirement is well-known. The diagram piece \( M^\mu \) acts, after all, like a source for the ‘photon’, and (since we work in the momentum representation) it is the Fourier transform of a spacetime current \( J^\mu \). The requirement of Eq. (158) is seen to be

\[
\partial_\mu J^\mu = 0,
\] (159)

which is nothing but conservation of electromagnetic charge in the case of photons.

Another way to arrive at the same result is to consider more closely the statement that ‘the electromagnetic field is transverse’. This is obviously not a Lorentz-invariant statement: even if the polarization vector happens to be transverse in one frame, a simple boost may make it non-transverse. In general, a massless spin-1 particle will have momentum and polarization given by

\[
p^\mu = (|\vec{p}|, \vec{\varepsilon})\ , \ \varepsilon^\mu = (\varepsilon^0, \vec{\varepsilon})\ , \ |\vec{p}|\varepsilon^0 = \vec{p} \cdot \vec{\varepsilon}.
\] (160)

Now we can split \( \varepsilon \) into a longitudinal part \( \varepsilon_L \) and a transverse part \( \varepsilon_T \):

\[
\varepsilon = \varepsilon_L + \varepsilon_T\ , \ \varepsilon_L = \frac{(\vec{p} \cdot \varepsilon) \vec{p}}{|\vec{p}|^2} = \varepsilon^0 \frac{\vec{p}}{|\vec{p}|},
\] (161)

so that

\[
\varepsilon^\mu = (\varepsilon^0, \varepsilon_L + \varepsilon_T) = (0, \varepsilon_T) + \frac{\varepsilon^0}{|\vec{p}|} p^\mu,
\] (162)

with the same result, namely that the transversality statement is effectively Lorentz-invariant provided \( M^\mu p_\mu = 0 \).

So far, we have been able to decouple external longitudinally polarized photons, but a danger lurks yet, since photons may of course also occur as internal lines. To solve this, we require that condition (158) also holds for off-shell momenta \( k^\mu \). Consider the diagram(s) given by

\[
\text{Diagram(s) given by}
\]

where we have indicated the momentum of the internal massless spin-1 line by \( k^\mu \). We can write this, in
the almost-massless case, as

\[ \mathcal{M} = \mathcal{M}_1^\mu \frac{i(-g_{\mu\nu} + k_\mu k_\nu/m^2)}{k^2 - m^2 + i\epsilon} \mathcal{M}_2^\nu. \]  

(163)

We now request that for all such diagrams,

\[ \mathcal{M}_1^\mu k_\mu = 0 \quad \text{and} \quad \mathcal{M}_2^\nu k_\nu = 0. \]  

(164)

This gauge invariance requirement will allow us to let \( m \) go to zero. The condition is too strict to satisfy for each individual diagram: the best we can arrange is that it holds for the complete set of diagrams that make up \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \). That, in turn, means that it is extremely dangerous to leave out diagrams, since delicate cancellations are necessary to satisfy gauge invariance: forgetting one term can lead to a cross section that is off by many orders of magnitude.

Finally, a word on the connection with Dirac particles. For those, the spin vector \( s^\mu \) blows up just like \( \epsilon^\mu \) for longitudinal polarization. It is one of the many small miracles of Dirac algebra that \( u(p, s) \) remains finite even if \( s^\mu \) explodes; our insistence on gauge invariance for spin-1 particles is necessary because no such mechanism is around for \( \epsilon^\mu \).
INTRODUCTION TO QCD

Michelangelo L. Mangano

CERN, TH Division, Geneva, Switzerland

Abstract
I review in this series of lectures the basics of perturbative quantum chromodynamics and some simple applications to the physics of high-energy collisions.

1. INTRODUCTION

Quantum Chromodynamics (QCD) is the theory of strong interactions. It is formulated in terms of elementary fields (quarks and gluons), whose interactions obey the principles of a relativistic QFT, with a non-abelian gauge invariance SU(3). The emergence of QCD as theory of strong interactions could be reviewed historically, analyzing the various experimental data and the theoretical ideas available in the years 1960–1973 (see e.g. Refs. [17,18]). To do this accurately and usefully would require more time than I have available. I therefore prefer to introduce QCD right away, and to use my time in exploring some of its consequences and applications. I will therefore assume that you all know more or less what QCD is! I assume you know that hadrons are made of quarks, that quarks are spin-1/2, colour-triplet fermions, interacting via the exchange of an octet of spin-1 gluons. I assume you know the concept of running couplings, asymptotic freedom and of confinement. I shall finally assume that you have some familiarity with the fundamental ideas and formalism of QED: Feynman rules, renormalization, gauge invariance.

If you go through lecture series on QCD (e.g., the lectures given in previous years at the CERN Summer School, Refs. [9–11]), you will hardly ever find the same item twice. This is because QCD today covers a huge set of subjects and each of us has his own concept of what to do with QCD and of what are the “fundamental” notions of QCD and its “fundamental” applications. As a result, you will find lecture series centred around non-perturbative applications, (lattice QCD, sum rules, chiral perturbation theory, heavy quark effective theory), around formal properties of the perturbative expansion (asymptotic behaviour, renormalons), techniques to evaluate complex classes of Feynman diagrams, or phenomenological applications of QCD to possibly very different sets of experimental data (structure functions, deep-inelastic scattering (DIS) sum rules, polarized DIS, small $x$ physics (including hard pomerons, diffraction), LEP physics, $p\bar{p}$ collisions, etc.

I can anticipate that I will not be able to cover or to simply mention all of this. After introducing some basic material, I will focus on some elementary applications of QCD in high-energy $e^+e^-$, $ep$ and $p\bar{p}$ collisions. The outline of these lectures is the following:

1. Gauge invariance and Feynman rules for QCD.
2. Renormalization, running coupling, renormalization group invariance.
3. QCD in $e^+e^-$ collisions: from quarks and gluons to hadrons, jets, shape variables.
4. QCD in lepton-hadron collisions: DIS, parton densities, parton evolution.
5. QCD in hadron-hadron collisions: formalism, $W/Z$ production, jet production.

Given the large number of papers which contributed to the development of the field, it is impossible to provide a fair bibliography. I therefore limit my list of references to some excellent books and review articles covering the material presented here, and more. Papers on specific items can be easily found by consulting the standard hep-th and hep-ph preprint archives.
2. QCD FEYNMAN RULES

There is no free lunch, so before starting with the applications, we need to spend some time developing the formalism and the necessary theoretical ideas. I will dedicate to this purpose the first two lectures. Today, I concentrate on Feynman rules. I will use an approach which is not canonical, namely it does not follow the standard path of the construction of a gauge invariant Lagrangian and the derivation of Feynman rules from it. I will rather start from QED, and empirically construct the extension to a non-Abelian theory by enforcing the desired symmetries directly on some specific scattering amplitudes. Hopefully, this will lead to a better insight into the relation between gauge invariance and Feynman rules. It will also provide you with a way of easily recalling or checking your rules when books are not around!

2.1. Summary of QED Feynman rules

We start by summarizing the familiar Feynman rules for Quantum Electrodynamics (QED). They are obtained from the Lagrangian:

\[ \mathcal{L} = \bar{\psi} (i\gamma \cdot \mathbf{D} - m) \psi - e^2 \bar{\psi} \gamma^\mu A_\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \]  

where \( \psi \) is the electron field, of mass \( m \) and coupling constant \( e \), and \( F^{\mu\nu} \) is the electromagnetic field strength.

\[ F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \]  

The resulting Feynman rules are summarized in the following table:

\[
\begin{align*}
\begin{array}{ccc}
\text{Particle} & \text{Symbol} & \text{Feynman rule} \\
\hline
\mu & \mu & \frac{i}{\not{p} - m + ie} \left( \frac{\not{p} + m}{p^2 - m^2 + ie} \right) (3) \\
\nu & \nu & -ie^{-1} g_{\mu\nu} (\text{Feynman gauge}) (4) \\
\gamma & \gamma & -ie\gamma_\mu Q (Q = -1 \text{ for the electron, } Q = 2/3 \text{ for the u-quark, etc.}) (5) \\
\hline
\end{array}
\end{align*}
\]

Let us start by considering a simple QED process, \( e^+e^- \rightarrow \gamma\gamma \) (for simplicity we shall always assume \( m = 0 \)):

\[
\begin{align*}
\begin{array}{ccc}
\text{Particle} & \text{Symbol} & \text{Feynman rule} \\
\hline
\mathbf{q} & \mu & -ie\gamma_\mu Q = D_1 + D_2 \\
\mathbf{k}_1 & \nu & \\
\mathbf{k}_2 & \nu & \\
\mathbf{q} & & \\
\end{array}
\end{align*}
\]

The total amplitude \( M_{\gamma} \) is given by:

\[
\frac{ie^2}{2} M_{\gamma} \equiv D_1 + D_2 = \bar{\nu}(\mathbf{q}) f_2 \left( \frac{1}{\mathbf{q} - \mathbf{k}_1} f_1 u(q) + \bar{\nu}(\mathbf{q}) f_1 \right) \frac{1}{\mathbf{q} - \mathbf{k}_2} f_2 n(q) \equiv M_{\mu\nu} \epsilon_1^{\mu} \epsilon_2^{\nu}. (7)
\]

Gauge invariance demands that

\[
\epsilon_2^{\nu} \partial^\mu M_{\mu\nu} = \epsilon_1^{\mu} \partial^\nu M_{\mu\nu} = 0. (8)
\]

\( M_{\mu} \equiv M_{\mu\nu} \epsilon_2^{\nu} \) is in fact the current that couples to the photon \( k_1 \). Charge conservation requires \( \partial_\mu M^\mu = 0 \):

\[
\partial_\mu M^\mu = 0 \Rightarrow \frac{d}{dt} \int M^0 d^3x = \int \partial_0 M^0 d^3x
\]

42
In momentum space, this means
\[ k_1^\mu M_{\mu} = 0. \] (10)
Another way of saying this is that the theory is invariant if \( \epsilon_\mu(k) \to \epsilon_\mu(k) + f(k) k_\mu \). This is the standard Abelian gauge invariance associated to the vector potential transformations:
\[ A_\mu(x) \to A_\mu(x) + \partial_\mu f(x) \] (11)

Let us verify that \( M_\gamma \) is indeed gauge invariant. Using \( \hat{g}_u (q) = \bar{v}(\bar{q}) \hat{g} = 0 \) from the Dirac equation, we can rewrite \( k_1^\mu M_{\mu} \) as:
\[
k_1^\mu \epsilon_2^\mu M_{\mu\nu} = \bar{v}(\bar{q}) \hat{f}_2 \frac{1}{\hat{g} - \gamma_1} (\hat{g}_1 - \hat{g}) u(q) + \bar{v}(\bar{q}) (\hat{g}_1 - \hat{q}) \frac{1}{\hat{k}_1 - \hat{q}} \hat{f}_2 u(q) \\
= -\bar{v}(\bar{q}) \hat{f}_2 u(q) + \bar{v}(\bar{q}) \hat{f}_2 u(q) = 0. \] (12)

Notice that the two diagrams are not individually gauge invariant, only the sum is. Notice also that the cancellation takes place independently of the choice of \( \epsilon_2 \). The amplitude is therefore gauge invariant even in the case of emission of non-transverse photons.

Let us try now to generalize our QED example to a theory where the “electrons” carry a non-Abelian charge, i.e., they transform under a non-trivial representation \( R \) of a non-Abelian group \( G \) (which, for the sake of simplicity, we shall always assume to be of the \( SU(N) \) type. Likewise, we shall refer to the non-abelian charge as “colour”). The standard current operator belongs to the product \( R \otimes \bar{R} \). The only representation that belongs to \( R \otimes \bar{R} \) for any \( R \) is the adjoint representation. Therefore the field that couples to the colour current must transform as the adjoint representation of the group \( G \).

So the only generalization of the photon field to the case of a non-Abelian symmetry is a set of vector fields transforming under the adjoint of \( G \), and the simplest generalization of the coupling to fermions takes the form:
\[
\begin{align*}
\nonumber \left( \begin{array}{ccc}
\lambda_a \\
\gamma_{ik} \\
\gamma_{mn}
\end{array} \right) & = i g \lambda^a_{ki} \gamma^\mu_{mn}, \\
\nonumber \left( \begin{array}{ccc}
\lambda_i \\
\gamma_{jn}
\end{array} \right) & = 0
\end{align*}
\] (13)

where the matrices \( \lambda^a \) represent the algebra of the group on the representation \( R \). By definition, they satisfy the algebra:
\[ [\lambda^a, \lambda^b] = i f^{abc} \lambda^c \] (14)
for a fixed set of structure constants \( f^{abc} \), which uniquely characterize the algebra. We shall call quarks \( (q) \) the fermion fields in \( R \) and gluons \( (g) \) the vector fields which couple to the quark colour current.

The non-abelian generalization of the \( e^+ e^- \to \gamma \gamma \) process is the \( q\bar{q} \to gg \) annihilation. Its amplitude can be evaluated by including the \( \lambda \) matrices in Eq. (6):
\[
\frac{i}{e^2} M_\gamma \to \frac{i}{g^2} M_g \equiv (\lambda^b \lambda^a)_{ij} D_1 + (\lambda^a \lambda^b)_{ij} D_2
\] (15)

with \((a, b)\) colour labels (i.e. group indices) of gluons 1 and 2, \((i, j)\) colour labels of \( \bar{q}, q \), respectively. Using Eq. (14), we can rewrite (15) as:
\[
M_g = (\lambda^a \lambda^b)_{ij} M_\gamma - g^2 f^{abc} \lambda^c_{ij} D_1.
\] (16)
If we want the charge associated with the group $G$ to be conserved, we still need to demand

$$k_1^\mu\epsilon_i^\nu \ M_g^{\mu\nu} = \epsilon_{1}^\mu k_2^\nu \ M_g^{\mu\nu} = 0.$$  \hspace{1cm} (17)

Substituting $\epsilon_1^\mu \rightarrow k_1^\mu$ in (16) we get instead, using (12):

$$k_1\mu M_g^{\mu} = -g^2 f^{abc} \lambda_i^a \ \bar{\psi}_i(q) \ \gamma_\mu \ u_i(q) .$$  \hspace{1cm} (18)

The gauge cancellation taking place in QED between the two diagrams is spoiled by the non-Abelian nature of the coupling of quarks to gluons (i.e., $\lambda^a$ and $\lambda^b$ do not commute, and $f^{abc} \neq 0$).

The only possible way to solve this problem is to include additional diagrams. That new interactions should exist is by itself a reasonable fact, since gluons are charged (i.e., they transform under the symmetry group) and might want to interact among themselves. If we rewrite (18) as follows:

$$k_1\mu M_g^{\mu} = i \ f^{abc} g_2 \epsilon_2^\mu \times \left( i g \lambda_i^a \ \bar{\psi}_i(q) \ \gamma_\mu \ u_i(q) \right) ,$$  \hspace{1cm} (19)

we can recognize in the second factor the structure of the $q\bar{q}g$ vertex. The first factor has the appropriate colour structure to describe a triple-gluon vertex, with $a, b, c$ the colour labels of the three gluons:

Equation (19) therefore suggests the existence of a coupling like (20), with a Lorentz structure $V_{\mu_1\mu_2\mu_3}$ to be specified, giving rise to the following contribution to $q\bar{q} \rightarrow gg$:

$$g f^{abc} V_{\mu_1\mu_2\mu_3}(k_1, k_2, k_3) \ .$$  \hspace{1cm} (20)

We now need to find $V_{\mu_1\mu_2\mu_3}(p_1, p_2, p_3)$ and to verify that the contribution of the new diagram to $k_1 \cdot M_g$ cancels that of the first two diagrams. We will now show that the constraints of Lorentz invariance, Bose symmetry and dimensional analysis uniquely fix $V$, up to an overall constant factor.

Dimensional analysis fixes the coupling to be linear in the gluon momenta. This is because each vector field carries dimension 1, there are three of them, and the interaction must have total dimension equal to 4. So at most one derivative (i.e. one power of momentum) can appear at the vertex. In principle, if some mass parameter were available, higher derivatives could be included, with the appropriate powers of the mass parameter appearing in the denominator. This is however not the case. It is important to remark that the absence of interactions with higher number of derivatives is also crucial for the renormalizability of the interaction.

Lorentz invariance requires then that $V$ be built out of terms of the form $g_{\mu_1\mu_2} p_3$. Bose symmetry requires $V$ to be fully antisymmetric under the exchange of any pair $(\mu_1, p_1) \leftrightarrow (\mu_2, p_2)$ since the colour structure $f^{abc}$ is totally antisymmetric. As a result, for example, a term like $g_{\mu_1\mu_2} p_3^{\mu_3}$ vanishes under antisymmetrization, while $g_{\mu_1\mu_2} p_1^{\mu_3}$ doesn’t. Starting from this last term, we can easily add the pieces required to obtain the full antisymmetry in all three indices. The result is unique, up to an overall factor:

$$V_{\mu_1\mu_2\mu_3} = \sqrt{\frac{\mu_1 g_{\mu_1\mu_2} + \mu_2 g_{\mu_2\mu_3} + \mu_3 g_{\mu_3\mu_1}}{6}} .$$  \hspace{1cm} (22)
To test the gauge variation of the contribution $D_3$, we set $\mu_3 = \mu, \epsilon_1 = k_1$ and $k_3 = -(k_1 + k_2)$ in Eq. (21), and we get:

$$k_1^\mu e_2^\nu V_{\mu_1 \mu_2 \mu_3} (k_1, k_2, k_3) = V_0 \{ -(k_1 + k_2)\mu (k_1 \cdot \epsilon_2) + 2(k_1 \cdot k_2) e_2^\mu - (k_2 \cdot \epsilon_2) k_1^\mu \} .$$

(23)

The gauge variation is therefore:

$$k_1 \cdot D_3 = g^2 f^{abc} \chi^c V_0 \left[ \bar{\psi}(q)f_2 u(q) - \frac{k_2 \cdot \epsilon_2}{2 k_1 k_2} \bar{\psi}(q) k_1 u(q) \right].$$

(24)

The first term cancels the gauge variation of $D_1 + D_2$ provided $V_0 = 1$, the second term vanishes for a physical gluon $k_2$, since in this case $k_2 \cdot \epsilon_2 = 0$. $D_1 + D_2 + D_3$ is therefore gauge invariant but, contrary to the case of QED, only for physical external on-shell gluons.

Having introduced a three-gluon coupling, we can induce processes involving only gluons, such as $gg \rightarrow gg$:

$$
\begin{aligned}
\text{[Diagram]} & + \quad \text{[Diagram]} + \quad \text{[Diagram]}
\end{aligned}
$$

(25)

Once more it is necessary to verify the gauge invariance of this amplitude. It turns out that one more diagram is required, induced by a four-gluon vertex. Lorentz invariance, Bose symmetry and dimensional analysis uniquely determine once again the structure of this vertex. The overall factor is fixed by gauge invariance. The resulting Feynman rule for the 4-gluon vertex is given in Fig. 1.

You can verify that the 3- and 4-gluon vertices we introduced above are exactly those which arise from the Yang–Mills Lagrangian:

$$L_{YM} = -\frac{1}{4} \sum_a \bar{F}_{\mu \nu}^a F^{a \mu \nu} \quad \text{with} \quad F_{\mu \nu}^a = \partial_\mu A_\nu^a - g f^{abc} A_\mu^b A_\nu^c .$$

(26)

It can be shown that the 3- and 4-gluon vertices we generated are all is needed to guarantee gauge invariance even for processes more complicated than those studied in the previous simple examples. In other words, no extra 5- or more gluon vertices have to be introduced to achieve the gauge invariance of higher-order amplitudes. At the tree level this is the consequence of dimensional analysis and of the locality of the couplings (no inverse powers of the momenta can appear in the Lagrangian). At the loop level, these conditions are supplemented by the renormalizability of the theory [3,7].

Before one can start calculating cross-sections, a technical subtlety that arises in QCD when squaring the amplitudes and summing over the polarization of external states needs to be discussed. Let us again start from the QED example. Let us focus, for example, on the sum over polarizations of photon $k_1$:

$$\sum_{\epsilon_1} |M|^2 = \left( \sum_{\epsilon_1} \epsilon_1^\mu \epsilon_1^\nu \right) M_\mu M_\nu .$$

(27)

The two independent physical polarizations of a photon with momentum $k = (k_0; 0, 0)$ are given by $\epsilon_{L,R}^\mu = (0; 1, \pm i, 0)/\sqrt{2}$. They satisfy the standard normalization properties:

$$\epsilon_L \cdot \epsilon_L^* = -1 = \epsilon_R \cdot \epsilon_R^*, \quad \epsilon_L \cdot \epsilon_R^* = 0 .$$

45
\[
\begin{align*}
\delta^{ab} \frac{-i g^\alpha \beta}{p^2 + i\epsilon} & \quad \text{\text{(Feynman gauge)}} \\
\delta^{ab} \frac{i}{p^2 + i\epsilon} & \\
\delta^{ik} \frac{i}{\not{p} - m + i\epsilon} & \\
& = g f^{abc} \left[ g^{\alpha \beta} (p - q)^\gamma + g^{\beta \gamma} (q - r)^\alpha + g^{\gamma \alpha} (r - p)^\beta \right] \\
& = -ig^2 f^{xac} f^{xbd} \left( g^{\alpha \beta} g^\gamma \delta - g^{\alpha \delta} g^\beta \gamma \right) \\
& -ig^2 f^{xad} f^{xbc} \left( g^{\alpha \beta} g^\gamma \delta - g^{\alpha \gamma} g^\beta \delta \right) \\
& -ig^2 f^{xab} f^{xcd} \left( g^{\alpha \gamma} g^\beta \delta - g^{\alpha \delta} g^\beta \gamma \right) \\
& = -g f^{abc} q^\alpha \\
& = ig \lambda^a_{ki} \gamma^\alpha_{mn}
\end{align*}
\]

Fig. 1: Feynman rules for QCD. The solid lines represent the fermions, the curly lines the gluons, and the dotted lines represent the ghosts.
We can write the sum over physical polarizations in a convenient form by introducing the vector \( \vec{k} = (k_0; 0, 0, -k_0) \):

\[
\sum_{i=L,R} \epsilon_i^\mu \epsilon_i^{\nu*} = \begin{pmatrix} 0 & \vec{0} \\ 1 & 0 & 0 \\ \vec{0} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = -g_{\mu\nu} + \frac{k_\mu \bar{k}_\nu + k_\nu \bar{k}_\mu}{k \cdot \bar{k}}. \tag{28}
\]

We could have written the sum over physical polarizations using any other momentum \( \ell_\mu \), provided \( k \cdot \ell \neq 0 \). This would be equivalent to a gauge transformation (prove it as an exercise). In QED the second term in Eq. (28) can be safely dropped, since \( k_\mu M^\mu = 0 \). As a cross check, notice that \( k_\mu M^\mu = 0 \) implies \( M_0 = M_3 \), and therefore:

\[
\sum_{i=L,R} |\epsilon_i \cdot M|^2 = |M_1|^2 + |M_2|^2 = |M_1|^2 + |M_2|^2 + |M_3|^2 - |M_0|^2 = -g^{\mu\nu} M_\mu M_\nu. \tag{29}
\]

Therefore, the production of the longitudinal and time-like components of the photon cancel each other. This is true regardless of whether additional external photons are physical or not, since the gauge invariance \( k_1 \cdot M = 0 \) shown in Eq. (12) holds regardless of the choice for \( \epsilon_2 \), as already remarked. In particular,

\[
k^{\mu_1}_1 k^{\mu_2}_2 M_{\mu_1 \mu_2} = 0 \tag{30}
\]

(for \( n \) photons, \( k^{\mu_1}_1 k^{\mu_2}_2 \ldots k^{\mu_n}_n M_{\mu_1 \ldots \mu_n} = 0 \) and the production of any number of unphysical photons vanishes. The situation in the case of gluon emission is different, since \( k_1 \cdot M \propto \epsilon_2 \cdot k_2 \), which vanishes only for a physical \( \epsilon_2 \). This implies that the production of one physical and one non-physical gluons is equal to 0, but the production of a pair of non-physical gluons is allowed! If \( \epsilon_2 \cdot k_2 \neq 0 \), then \( M_0 \) is not equal to \( M_3 \), and Eq. (28) is not equivalent to \( \sum \epsilon_\mu \epsilon_\nu^{*} = -g_{\mu\nu} \).

**Exercise:** show that

\[
\sum_{\text{non-physical}} |\epsilon_1^\mu \epsilon_2^\nu M_{\mu\nu}|^2 = \left| i g^2 f^{abc} \chi^c \frac{1}{2k_1 k_2} \bar{v}(q)k_\| u(q) \right|^2. \tag{31}
\]

In the case of non-Abelian theories, it is therefore important to restrict the sum over polarizations and (because of unitarity) the off-shell propagators to physical degrees of freedom with the choice of physical gauges. Alternatively, one has to undertake a study of the implications of gauge-fixing in non-physical gauges for the quantization of the theory (see Refs. [3,7]). The outcome of this analysis is the appearance of two colour-octet scalar degrees of freedom (called ghosts) whose rôle is to enforce unitarity in non-physical gauges. They will appear in internal closed loops, or will be pair-produced in final states. They only couple to gluons. Their Feynman rules are supplemented by the prescription that each closed loop should come with a \(-1\) sign, as if they obeyed Fermi statistics. Being scalars, this prescription breaks the spin-statistics relation, and leads as a result to the possibility that production probabilities be negative. This is precisely what is required to cancel the contributions of non-transverse degrees of freedom appearing in non-physical gauges. Adding the ghosts contribution to \( q\bar{q} \rightarrow gg \) decays (using the Feynman rules from Fig. 1) gives in fact

\[
\left| \begin{array}{c}
\gamma \\gamma
\end{array} \right|^2 = - \left| i g^2 f^{abc} \chi^c \frac{1}{2k_1 k_2} \bar{v}(q)k_\| u(q) \right|^2, \tag{32}
\]

which exactly cancels the contribution of non-transverse gluons in the non-physical gauge \( \sum \epsilon_\mu \epsilon_\nu^{*} = -g_{\mu\nu} \), given in Eq. 31.
The detailed derivation of the need for and properties of ghosts (including their Feynman rules and the “−1” prescription for loops) can be found in the suggested textbooks. I will not derive these results here since we will not need them for our applications (we will use physical gauges or will consider processes not involving the 3$g$ vertex). The full set of Feynman rules for the QCD Lagrangian is given in Fig. 1.

2.2. Some useful results in colour algebra

The presence of colour factors in the Feynman rules makes it necessary to develop some technology to evaluate the colour coefficients which multiply our Feynman diagrams. To be specific, we shall assume the gauge group is $SU(N)$. The fundamental relation of the algebra is

$$[\lambda^a, \lambda^b] = if^{abc}\lambda^c,$$

with $f^{abc}$ totally antisymmetric. This relation implies that all $\lambda$ matrices are traceless. For practical calculations, since we will always sum over initial, final, and intermediate state colours, we will never need the explicit values of $f^{abc}$. All of the results can be expressed in terms of group invariants (a.k.a. Casimirs), some of which we will now introduce. The first such invariant ($T_F$) is chosen to fix the normalization of the matrices $\lambda$:

$$\text{tr}(\lambda^a\lambda^b) = T_F\delta_{ab},$$

where by convention $T_F = 1/2$ for the fundamental representation. Should you change this convention, you would need to change the definition (i.e. the numerical value) of the coupling constant $g$, since $g\lambda^a$ appears in the Lagrangian and in the Feynman rules.

**Exercise**: Show that $\text{tr}(\lambda^a\lambda^b)$ is indeed a group invariant. Hint: write the action on $\lambda^a$ of a general group transformation with infinitesimal parameters $\epsilon^b$ as follows:

$$\delta\lambda^a = \sum_{b,c} \epsilon_b f^{abc}\lambda^c.$$  

The definition of $T_F$ allows to evaluate the colour factor for an interesting diagram, i.e. the quark self-energy:

$$i\lambda^a \bar{\psi} x \frac{1}{2\pi} \lambda^a \bar{\psi} y \sim \sum_{i,j} (\lambda^a\lambda^a)_{ij} = C_F\delta_{ij}.$$  

The value of $C_F$ can be obtained by tracing the relation above:

$$C_FN = \text{tr} \sum_a \lambda^a\lambda^a = \delta^{ab} T_F\delta_{ab} = \frac{N^2 - 1}{2},$$

where we used the fact that $\delta^{ab}\delta_{ab} = N^2 - 1$, the number of matrices $\lambda^a$ (and of gluons) for $SU(N)$.

There are some useful graphical tricks (which I learned from P. Nason, Ref. [9]) which can be used to evaluate complicated expressions. The starting point is the following representation for the quark and gluon propagators, and for the $q\bar{q}g$ and $ggg$ interaction vertices:

```
fermion
---

gluon
---
```
\[
\frac{1}{\sqrt{2}} \left( \begin{array}{c}
\text{Fermion-Gluon Vertex } (t^a) \\
\end{array} \right)  \\
\frac{1}{\sqrt{2}} \left( \begin{array}{c}
\text{3-Gluon Vertex } (f^{abc}) \\
\end{array} \right)
\]

Contraction over colour indices is obtained by connecting the respective colour (or anticolour) lines. A closed loop of a colour line gives rise to a factor \(N\), since the closed loop is equivalent to the trace of the unit matrix. So the above representation of the \(q\bar{q}g\) vertex embodies the idea of “colour conservation”, whereby the colour-anticolour quantum numbers carried by the \(q\bar{q}\) pair are transferred to the gluon. The piece proportional to \(1/N\) in the \(q\bar{q}g\) vertex appears only when the colour of the quark and of the antiquark are the same. It ensures that \(\lambda^a\) is traceless, as it should. This can be easily checked as an exercise. The factor \(1/\sqrt{2}\) is related to the chosen normalization of \(T_F\).

As a first example of applications, let us reevaluate \(C_F\):

\[
\frac{1}{\sqrt{2}} \left( \begin{array}{c}
\end{array} \right) \times \frac{1}{\sqrt{2}} \left( \begin{array}{c}
\end{array} \right) = \frac{1}{2} \left( \begin{array}{c}
\end{array} \right) + \frac{1}{N^2} \left( \begin{array}{c}
\end{array} \right) = \delta_{ij} \frac{N^2 - 1}{2N}. 
\]

As an exercise, you can calculate the colour factor for \(q\bar{q} \rightarrow q\bar{q}\) scattering, and show that:

\[
\sum_a (\lambda^a)_{ij} (\lambda^a)_{\ell k} = \frac{1}{2} \left( \begin{array}{c}
\end{array} \right) = \frac{1}{2} \left( \begin{array}{c}
\end{array} \right) = \frac{1}{2} \left( \delta_{ik} \delta_{\ell j} - \frac{1}{N} \delta_{ij} \delta_{\ell k} \right). 
\]

This result can be used to evaluate the one-loop colour factors for the interaction vertex with a photon:

\[
\frac{1}{2} \left( \begin{array}{c}
\end{array} \right) = \frac{1}{2} \frac{N^2 - 1}{N} \delta_{ij} = C_F \delta_{ij}. 
\]
For the interaction with a gluon we have instead:

\[
\begin{align*}
\frac{1}{\sqrt{2}} & \left( \begin{array}{c}
\frac{1}{N} \\
0 \\
-\frac{1}{N}
\end{array} \right) \times \frac{1}{2} \left( \begin{array}{c}
0 \\
-\frac{1}{N} \\
\frac{1}{N} + \frac{1}{N^2}
\end{array} \right) \\
& = \frac{1}{2\sqrt{2}} \left( \begin{array}{c}
\frac{1}{N} \\
0 \\
-\frac{1}{N}
\end{array} \right) \\
& = -\frac{1}{2N} \sqrt{2} \left( \begin{array}{c}
\frac{1}{N} \\
0 \\
-\frac{1}{N}
\end{array} \right) = -\frac{1}{2N}
\end{align*}
\]

(45)

Notice that in the case of the coupling to the photon the \( q\bar{q} \) pair is in a colour-singlet state. The gluon exchange effect in this case has a positive sign (\( \Rightarrow \) attraction). In the case of the coupling to the gluon the \( q\bar{q} \) pair is in a colour-octet state, and the gluon-exchange correction has a negative sign relative to the Born interaction. The force between a \( q\bar{q} \) pair is therefore attractive if the pair is in a colour-singlet, while it is repulsive if it is in a colour-octet state! This gives a qualitative argument for why no colour-octet \( q\bar{q} \) bound state exists.

The remaining important relation that one needs is the following:

\[
\sum_{a,b} f^{abc} f^{abd} = C_A \delta^{cd} \quad \text{with} \quad C_A = N.
\]

(46)

You can easily prove it by using the graphical representation given in Eq. (41), or by using Eq. (43) and \( f^{abc} = -2i \text{tr}([\lambda^a, \lambda^b] \lambda^c) \).

3. RENORMALIZATION, OR: “THEORISTS ARE NOT AFRAID OF INFINITIES!”

QCD calculations are extremely demanding. Although perturbative, the size of the coupling constant even at rather large values of the exchanged momentum, \( Q^2 \), is such that the convergence of the perturbative expansion is slow. Several orders of perturbation theory (PT) are required in order to obtain a good accuracy. The complexity of the calculations grows dramatically with the order of the approximation. As an additional complication, the evaluation of a large class of higher-order diagrams gives rise to results which are a priori ill-defined, namely to infinities. A typical example of what is known as an ultraviolet divergence, appears when considering the corrections to the quark self-energy. Using the Feynman rules presented in the previous lecture, one can obtain:

\[
\begin{align*}
\frac{1}{p-1} \int \frac{d^4 \ell}{(2\pi)^4} & \gamma_\mu \frac{i}{p+\ell} \gamma_\nu \left( -\frac{i g^{\mu\nu}}{\ell^2} \right) \equiv i \phi \Sigma(p),
\end{align*}
\]

(47)

where simple manipulations lead to the following expression for \( \Sigma(p) \):

\[
\Sigma(p) = i C_F \int \frac{d^4 \ell}{(2\pi)^4} \frac{1}{\ell^2 (p+\ell)^2},
\]

(48)
which is logarithmically divergent in the ultraviolet ($|\ell| \to \infty$) region. In this lecture we will discuss how to deal with these infinities. To start with, we study a simple example taken from standard electrostatics.

### 3.1. The potential of an infinite line of charge

Let us consider a wire of infinite length, carrying a constant charge density $\lambda$. By definition, the dimensions of $\lambda$ are $[\text{length}]^{-1}$. Our goal is to evaluate the electric potential, and eventually the electric field, in a point $P$ at distance $R$ from the wire. There is no need to do any calculation to anticipate that the evaluation of the electric potential will cause some problem. Using the fact that the potential should be linear in the charge density $\lambda$, we write $V(R) = \lambda f(R)$. Since the potential itself has the dimensions of $[\text{length}]^{-1}$, we clearly see that there is no room for $f(R)$ to have any non-trivial functional dependence on $R$. The problem is made explicit if we try to evaluate $V(R)$ using the standard EM formulas:

$$V(R) = \int \frac{\lambda(r)}{r} \, dx = \lambda \int_{-\infty}^{+\infty} \frac{dx}{\sqrt{R^2 + x^2}} , \quad (49)$$

where the integral runs over the position $x$ on the wire. This integral is logarithmically divergent, and the potential is ill-defined. We know however that this is not a serious issue, since the potential itself is not a physical observable, only the electric field is measurable. Since the electric field is obtained by taking the gradient of the scalar potential, it will be proportional to

$$\vec{E}(R) \sim \lambda \int_{-\infty}^{+\infty} \frac{dx}{(R^2 + x^2)^{3/2}} , \quad (50)$$

which is perfectly convergent. It is however interesting to explore the possibility of providing a useful operative meaning to the definition of the scalar potential. To do that, we start by regularizing the integral in Eq. (49). This can be done by introducing the regularized $V(R)$ defined as:

$$V_{\Lambda}(R) = \int_{-\Lambda}^{\Lambda} \frac{dx}{\sqrt{R^2 + x^2}} = \lambda \log \left[ \frac{\sqrt{\Lambda^2 + R^2} + \Lambda}{\sqrt{\Lambda^2 + R^2} - \Lambda} \right] . \quad (51)$$

We can then define the electric field as

$$\vec{E}(R) = \lim_{\Lambda \to \infty} [-\vec{\nabla} V_{\Lambda}(R)] . \quad (52)$$

It is easy to check that this prescription leads to the right result:

$$\vec{E}(R) = \lim_{\Lambda \to \infty} \frac{2\Lambda}{R} \frac{\Lambda}{\sqrt{\Lambda^2 + R^2}} \to \frac{2\lambda}{R} \hat{R} . \quad (53)$$

Notice that in this process we had to introduce a new variable $\Lambda$ with the dimension of a length. This allows us to solve the puzzle first pointed out at the beginning. At the end, however, the dependence of the physical observable (i.e. the electric field) on this extra parameter disappears. Notice also that the object:

$$\delta V = \lim_{\Lambda \to \infty} [V_{\Lambda}(r_2) - V_{\Lambda}(r_1)] = \lambda \log \left( \frac{r_1^2}{r_2^2} \right) \quad (54)$$

is well defined. This suggests a way of defining the potential which is meaningful even in the $\Lambda \to \infty$ limit. We can renormalize the potential, by subtracting $V(R)$ at some fixed value of $R = R_0$, and taking the $\Lambda \to \infty$ limit:

$$V(R) \to V(R) - V(R_0) = \lambda \log \left( \frac{R_0}{R^2} \right) . \quad (55)$$


The non-physical infinities present in $V(R)$ and $V(R_0)$ cancel each other, leaving a finite result, with a non-trivial $R$-dependence. Once again, this is possible because a dimensionful parameter (in this case $R_0$) has been introduced.

This example suggests a strategy for dealing with divergencies:

1. Identify an appropriate way to regularize infinite integrals
2. Absorb the divergent terms into a redefinition of fields or parameters, e.g., via subtractions. This step is usually called renormalization.
3. Make sure the procedure is consistent, by checking that the physical results do not depend on the regularization prescription.

In the rest of this lecture I will explain how this strategy is applied to the case of ultraviolet divergencies encountered in perturbation theory.

### 3.2. Dimensional regularization

The typical expressions we have to deal with have the form:

$$I(M^2) = \int \frac{d^4\ell}{(2\pi)^4} \frac{1}{(\ell^2 + M^2)^2}.$$  

(56)

You can easily show that the integral encountered in the quark self-energy diagram can be rewritten as:

$$\frac{1}{\ell^2} \frac{1}{(\ell - p)^2} = \int_0^1 dx \frac{1}{(L^2 + M^2)^2}, \text{ with } L = \ell - xp, M^2 = x(1-x)p^2.$$  

(57)

The most straightforward extension of the ideas presented above in the case of the infinite charged wire is to regularize the integral using a momentum cutoff, and to renormalize it with a subtraction (for example, $I(M^2) - I(M_0^2)$). Experience has shown, however, that the best way to regularize $I(M^2)$ is to take the analytic continuation of the integral in the number of space-time dimensions. In fact

$$I_D(M^2) = \int \frac{d^D\ell}{(2\pi)^D} \frac{1}{(\ell^2 + M^2)^2}$$  

(58)

is finite $\forall D < 4$. If we could assign a formal meaning to $I_D(M^2)$ for continuous values of $D$ away from $D = 4$, we could then perform all our manipulations in $D \neq 4$, regulate the divergences, renormalize fields and couplings, and then go back to $D = 4$.

To proceed, one defines (for Euclidean metrics):

$$d^D\ell = d\Omega_{D-1} \ell^{D-1}d\ell,$$  

(59)

with $d\Omega_{D-1}$ the differential solid angle in $D$ dimensions. $\Omega_{D-1}$ is the surface of a $D$-dimensional sphere. It can be obtained by using the following formal identity:

$$\int d^D\ell \ e^{-\ell^2} = \left[ \int d\ell \ e^{-\ell^2} \right]^D = \pi^{D/2}.$$  

(60)

The integral can also be evaluated, using Eq. (59), as

$$\int d^D\ell \ e^{-\ell^2} = \Omega_D \int_0^\infty \ell^{D-1} e^{-\ell^2} d\ell = \Omega_D \frac{1}{2} \int_0^\infty \ell^{D-2} e^{-\ell^2} \ell^2 d\ell$$

$$= \Omega_D \frac{1}{2} \int_0^\infty dx \ e^{-x} x^{D-2} \equiv \Omega_D \frac{1}{2} \Gamma \left( \frac{D}{2} \right).$$  

(61)
Comparing Eqs. (60) and (61) we get:

\[ I_D(M^2) = \frac{1}{(4\pi)^{D/2}} \frac{1}{\Gamma(D/2)} \int_0^\infty dx x^{D-2} (x + M^2)^2 = \frac{1}{(4\pi)^{D/2}} \frac{\Gamma(2 - D/2)}{\Gamma(2)} (M^2)^{D/2 - 2} \quad (62) \]

Defining \( D = 4 - 2\epsilon \) (with the understanding that \( \epsilon \) will be taken to 0 at the end of the day), and using the small-\( \epsilon \) expansion:

\[ \Gamma(\epsilon) = \frac{1}{\epsilon} - \gamma_\epsilon + \mathcal{O}(\epsilon) \quad (63) \]

we finally obtain:

\[ (4\pi)^2 I_D(M^2) \rightarrow \frac{1}{\epsilon} - \log 4\pi M^2 - \gamma_\epsilon \quad (64) \]

The divergent part of the integral is then regularized as a pole in \((D - 4)\). The \( M \)-dependent part of the integral behaves logarithmically, as expected because the integral itself was dimensionless in \( D = 4 \). The \( 1/\epsilon \) pole can be removed by a subtraction:

\[ I(M^2) = I(\mu^2) + (4\pi)^2 \log \left( \frac{\mu^2}{M^2} \right) \quad (65) \]

where the subtraction scale \( \mu^2 \) is usually referred to as the “renormalization scale”.

One can prove (and you will find this in the quoted textbooks) that other divergent integrals which appear in other loop diagrams can be regularized in a similar fashion, with the appearance of \( 1/\epsilon \) poles. Explicit calculations and more details on this technique can be found in the literature quoted at the end.

### 3.3. Renormalization

Let us come back now to our quark self-energy diagram, Eq. (47). After regulating the divergence using dimensional regularization, we can eliminate it by adding a counterterm to the Lagrangian:

\[ \mathcal{L} \rightarrow \mathcal{L} + \Sigma(p) \bar{\psi} i\partial \psi = [1 + \Sigma(p)] \bar{\psi} i\partial \psi + \ldots \quad (66) \]

In this way, the corrections at \( O(g^2) \) to the inverse propagator are finite:

\[ \begin{array}{c}
\text{---} \\
+ \quad \Sigma(p)
\end{array} = -i\slashed{p}\Sigma(p) + i\slashed{p}\Sigma(p) = 0 \quad (67) \]

The inclusion of this counterterm can be interpreted as a renormalization of the quark wave function. To see this, it is sufficient to define:

\[ \psi_R = \left( 1 + \Sigma(p^2) \right)^{1/2} \psi \quad (68) \]

and verify that the kinetic part of the Lagrangian written in terms of \( \psi_R \) takes again the canonical form.

It may seem that this regularization/renormalization procedure can always be carried out, with all possible infinities being removed by ad hoc counterterms. This is not true. That these subtractions can be performed consistently for any possible type of divergence which develops in PT is a highly non-trivial fact. To convince you of this, consider the following example.
Let us study the QCD corrections to the interaction of quarks with a photon:

\[
\begin{align*}
\langle q | \bar{q} \Gamma \gamma^\mu \gamma^5 | p \rangle &= (-ig)^2 C_F \int \frac{d^4 \ell}{(2\pi)^4} \left[ \gamma^\mu \frac{i}{\ell + \frac{\ell}{2}} \left( -i e \gamma^\mu \right) \frac{i}{\ell + \frac{\ell}{2}} \gamma^5 \right] \left( \frac{-i}{\ell^2} \right) \\
&= -ig^2 C_F \int \frac{d^4 \ell}{(2\pi)^4} (-2)(\ell + f)\Gamma^\mu(\ell + f) \frac{1}{\ell^2(p + \ell)^2(\ell + \ell)^2}
\end{align*}
\]

leading div. \(-ig^2(-2)C_F \int \frac{d^4 \ell}{(2\pi)^4} \frac{f\Gamma^\mu f}{\ell^2(p + \ell)^2(\ell + \ell)^2} \right) \quad \text{def} = i e \gamma^\mu V(q^2) .

It is easily recognized that \(V(q^2)\) is divergent. The divergence can be removed by adding a counterterm to the bare Lagrangian:

\[
\mathcal{L}_{\text{int}} = -e A_\mu \bar{\psi} \gamma^\mu \psi \rightarrow -e A_\mu \bar{\psi} \gamma^\mu \psi - eV(q^2)A_\mu \bar{\psi} \gamma^\mu \psi = -[1 + V(q^2)] e A_\mu \bar{\psi} \gamma^\mu \psi . \quad (69)
\]

If we take into account the counterterm that was introduced to renormalize the quark self-energy, the part of the quark Lagrangian describing the interaction with photons is now:

\[
\mathcal{L}_{q,\gamma} = \left[ 1 + \Sigma(p^2) \right] \bar{\psi} i \not{\partial} \psi - \left[ 1 + V(q^2) \right] e A_\mu \bar{\psi} \gamma^\mu \psi . \quad (70)
\]

Defining a renormalized charge by:

\[
e_R = \frac{1 + V(p^2)}{1 + \Sigma(q^2)} , \quad (71)
\]

we are left with the renormalized Lagrangian:

\[
\mathcal{L}_R = \bar{\psi}_R i \not{\partial} \psi_R + e_R A_\mu \bar{\psi}_R \gamma^\mu \psi_R . \quad (72)
\]

Can we blindly accept this result, regardless of the values of the counterterms \(V(p^2)\) and \(\Sigma(q^2)\)? The answer to this question is NO! Charge conservation, in fact, requires \(e_R = e\). The electric charge carried by a quark cannot be affected by the QCD corrections, and cannot be affected by the renormalization of QCD-induced divergencies. There are many ways to see that if \(e_R \neq e\) the electric charge would not be conserved in strong interactions. The simplest way is to consider the process \(e^+ \nu_e \rightarrow W^+ \rightarrow ud\). The electric charge of the initial state is +1 in units of \(e\). After including QCD corrections (which in the case of the interaction with a \(W^+\) are the same as those for the interaction of quarks with a photon), the charge of the final state is +1 in units of \(e_R\). Unless \(e_R = e\), the total electric charge would not be conserved in this process! The non-renormalization of the electric charge in presence of strong interactions is the fact that makes the charge of the proton to equal the sum of the charges of its constituent quarks, in spite of the complex QCD dynamics that holds the quarks together.

As a result, the renormalization procedure is consistent with charge conservation if and only if

\[
\frac{V(q^2)}{\Sigma(p^2)} \xrightarrow[q^2 \rightarrow 0]{} 1 . \quad (73)
\]

This identity should hold at all orders of perturbation theory. It represents a fundamental constraint on the consistency of the theory, and shows that the removal of infinities, by itself, is not a trivial trick which can be applied to arbitrary theories. Fortunately, the previous identity can be shown to hold. You can prove it explicitly at the one-loop order by explicitly evaluating the integrals defining \(V(q)\) and \(\Sigma(p)\).
To carry out the renormalization program for QCD at 1-loop order, several other diagrams in addition to the quark self-energy need to be evaluated. One needs the corrections to the gluon self-energy, to the coupling of a quark pair to a gluon, and to the 3-gluon coupling. Each of these corrections gives rise to infinities, which can be regulated in dimensional regularization. For the purposes of renormalization, it is useful to apply the concept of \( D \) dimensions not only to the evaluation of the infinite integrals, but to the full theory as well. In other words we should consider the Lagrangian as describing the interactions of fields in \( D \)-dimensions. Nothing changes in its form, but the canonical dimensions of fields and couplings will be shifted. This is because the action (defined as the integral over space-time of the lagrangian), is a dimensionless quantity. As a result, the canonical dimensions of the fields, and of the coupling constants, have to depend on \( D \):

\[
\int d^D x \mathcal{L}(x) = 0 \Rightarrow [\mathcal{L}] = D = 4 - 2\epsilon ,
\]

\[
[\partial_\mu \phi \partial^\mu \phi] = D \Rightarrow [\phi] = 1 - \epsilon ,
\]

\[
[\bar{\psi} \partial \psi] = D \Rightarrow [\psi] = 3/2 - \epsilon ,
\]

\[
[\bar{\psi} \gamma_5 A \psi] = D \Rightarrow [g] = \epsilon .
\]

The gauge coupling constant acquires dimensions! This is a prelude to the non-trivial behaviour of the renormalized coupling constant as a function of the energy scale ("running"). But before we come to this, let us go back to the calculation of the counter-terms and the construction of the renormalized Lagrangian.

Replace the bare fields and couplings with renormalized ones\(^1\):

\[
\psi_{\text{bare}} = Z_{1}^{1/2} \psi_R ,
\]

\[
A_{\mu \text{bare}} = Z_{3}^{1/2} A_{\mu R} ,
\]

\[
g_{\text{bare}} = Z_{g} \mu^\epsilon g_R .
\]

We explicitly extracted the dimensions out of \( g_{\text{bare}} \), introducing the dimensional parameter \( \mu \) (renormalization scale). In this way the renormalized coupling \( g_R \) is dimensionless (as it should be once we go back to 4-dimensions).

The Lagrangian, written in terms of renormalized quantities, becomes:

\[
\mathcal{L} = Z_2 \bar{\psi} i \partial \psi - \frac{1}{4} Z_3 F_{\mu \nu}^a F_{\alpha \beta}^a + Z_2 Z_3^{1/2} \mu^\epsilon \bar{\psi} A \psi + \text{(gauge fixing, ghosts, \ldots)} \quad (74)
\]

It is customary to define

\[
Z_1 = Z_g Z_2 Z_3^{1/2} . \quad (75)
\]

If we set \( Z_n = 1 + \delta_n \), we then obtain:

\[
\mathcal{L} = \bar{\psi} i \partial \psi - \frac{1}{4} F_{\mu \nu}^a F_{\alpha \beta}^a + \mu^\epsilon \bar{\psi} A \psi + \text{(ghosts, GM)}
\]

\[
+ \delta_2 \bar{\psi} i \partial \psi - \frac{1}{4} \delta_3 F_{\mu \nu}^a F_{\alpha \beta}^a + \delta_1 \mu^\epsilon \bar{\psi} A \psi . \quad (76)
\]

The counter-terms \( \delta_i \) are fixed by requiring the 1-loop Green functions to be finite. The explicit evaluation, which you can find carried out in detail, for example, in Refs. [7,3], gives:

---

\(^1\)For the sake of simplicity, here and in the following we shall assume the quarks as massless. The inclusion of the mass terms does not add any interesting new feature in what follows.
quark self-energy \quad \Rightarrow \quad \delta_2 = -C_F \left( \frac{\alpha_s}{4\pi} \frac{1}{\epsilon} \right), \quad (77)

gluon self-energy \quad \Rightarrow \quad \delta_3 = \left( \frac{5}{3} C_A - \frac{4}{3} n_f T_F \right) \alpha_s \frac{1}{4\pi \epsilon}, \quad (78)

q\bar{q}g \quad \text{vertex corrections} \quad \Rightarrow \quad \delta_1 = -(C_A + C_F) \frac{\alpha_s}{4\pi} \frac{1}{\epsilon}. \quad (79)

As usual we introduced the notation \( \alpha_s = g^2 / 4\pi \). The strong-coupling renormalization constant \( Z_g \) can be obtained using these results and Eq. (75):

\[
Z_g = \frac{Z_1}{Z_2 Z_3^{1/2}} = 1 + \delta_1 - \delta_2 - \frac{1}{2} \delta_3 = 1 + \frac{\alpha_s}{4\pi} \frac{1}{\epsilon} \left[ \frac{11}{6} C_A + \frac{2}{3} n_f T_F \right] \equiv 1 - \frac{1}{\epsilon} \left( \frac{b_0}{2} \right) \alpha_s. \quad (80)
\]

Notice the cancellation of the terms proportional to \( C_F \), between the quark self-energy \( (Z_2) \) and the abelian part of the vertex correction \( (Z_1) \). This is the same as in the case of the QCD non-renormalization of the electric coupling, discussed at the beginning of the lecture. The non-abelian part of the vertex correction contributes viceversa to the QCD coupling renormalization. This is a consequence of gauge invariance. The separation of the non-abelian contributions to the self-energy and to the vertex is not gauge-invariant, only their sum is. Notice also that the consistency of the renormalization procedure requires that the renormalized strong coupling \( g \) defining the strength of the interaction of quarks and gluons should be the same as that defining the interaction of gluons among themselves. If this didn’t happen, the gauge invariance of the \( q\bar{q} \to gg \) process so painfully achieved in the first lecture by fixing the coefficient of the 3-gluon coupling would not hold anymore at 1-loop! Once again, this additional constraint can be shown to hold through an explicit calculation.

### 3.4. Running of \( \alpha_s \)

The running of \( \alpha_s \) is a consequence of the renormalization-scale independence of the renormalization process. The bare coupling \( g_{bare} \) knows nothing about our choice of \( \mu \). The parameter \( \mu \) is an artifact of the regularization prescription, introduced to define the dimensionful coupling in \( D \) dimensions, and should not enter in measurable quantities. As a result:

\[
\frac{dg_{bare}}{d\mu} = 0. \quad (81)
\]

Using the definition of \( g \): \( g_{bare} = \mu^\epsilon Z_g \ g \), we then get

\[
\epsilon \mu^{2\epsilon} Z_g^2 \alpha_s + \mu^{2\epsilon} \alpha_s 2 \frac{dZ_g}{dt} + \mu^{2\epsilon} Z_g^2 \frac{d\alpha_s}{dt} = 0, \quad (82)
\]

where

\[
\frac{d}{dt} = \mu^2 \frac{d}{d\mu^2} = \frac{d}{d\log \mu^2}. \quad (83)
\]

\( Z_g \) depends upon \( \mu \) only via the presence of \( \alpha_s \). If we define

\[
\beta(\alpha_s) = \frac{d\alpha_s}{dt}, \quad (84)
\]

we then get:

\[
\beta(\alpha_s) + 2 \frac{\alpha_s}{Z_g} \frac{dZ_g}{d\alpha_s} \beta(\alpha_s) = -\epsilon \alpha_s. \quad (85)
\]
Using Eq. (80) and expanding in powers of $\alpha_s$, we get:

$$
\beta(\alpha_s) = \frac{-\epsilon \alpha_s}{1 + 2 \frac{Z_g}{Z} \frac{dZ_g}{d\alpha_s}} = \frac{-\epsilon \alpha_s}{1 - b_0 \alpha_s} = -b_0 \alpha_s^2 + O(\alpha_s^2, \epsilon),
$$

(86)

and finally:

$$
\beta(\alpha_s) = -b_0 \alpha_s^2 \quad \text{with} \quad b_0 = \frac{1}{2\pi} \left( \frac{11}{6} C_A - \frac{2}{3} n_f T_F \right)^N \frac{1}{12\pi} (33 - 2n_f).
$$

(87)

We can now solve Eq. (84), assuming $b_0 > 0$ (which is true provided the number of quark flavours is less than 16) and get the famous running of $\alpha_s$:

$$
\alpha_s(\mu^2) = \frac{1}{b_0 \log(\mu^2/\Lambda^2)}.
$$

(88)

The parameter $\Lambda$ describes the boundary condition of the first order differential equation defining the running of $\alpha_s$, and corresponds to the scale at which the coupling becomes infinity.

### 3.5. Renormalization group invariance

The fact that the coupling constant $\alpha_s$ depends on the unphysical renormalization scale $\mu$ should not be a source of worry. This is because the coupling constant itself is not an observable. What we observe are decay rates, spectra, or cross sections. These are given by the product of the coupling constant times some matrix element, which in general will acquire a non-trivial renormalization-scale dependence through the renormalization procedure. We therefore just need to check that the scale dependence of the coupling constant and of the matrix elements cancel each other, leaving results which do not depend on $\mu$.

Consider now a physical observable, for example the ratio $R = \sigma(e^+e^- \to \text{hadrons})/\sigma(e^+e^- \to \mu^+\mu^-)$. $R$ can be calculated in perturbation theory within QCD, giving rise to an expansion in the renormalized coupling $\alpha_s(\mu)$:

$$
R[\alpha_s, s/\mu^2] = 1 + \alpha_s f_1(t) + \alpha_s^2 f_2(t) + \ldots = \sum_{n=0}^{\infty} \alpha_s^n f_{(n)}(t),
$$

(89)

where $t = s/\mu^2$ (and we omitted a trivial overall factor $3\sum_f Q_f^2$). $R$ depends on $\mu$ explicitly via the functions $f_{(n)}(t)$ and implicitly through $\alpha_s$. Since $R$ is an observable, it should be independent of $\mu$, and the functions $f_{(n)}(t)$ cannot be totally arbitrary. In particular, one should have:

$$
\mu^2 \frac{dR}{d\mu^2} = 0 = \left[ \mu^2 \frac{\partial}{\partial \mu^2} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s} \right] R[\alpha_s, s/\mu^2] = 0.
$$

(90)

Before we give the general, formal solution to this differential equation, it is instructive to work out directly its form within perturbation theory.

$$
\mu^2 \frac{dR}{d\mu^2} = 0 = \beta(\alpha_s) f_1(t) + \alpha_s \mu^2 \frac{df_1}{d\mu^2} + 2\alpha_s \beta(\alpha_s) f_2(t) + \alpha_s^2 \mu^2 \frac{df_2}{d\mu^2} + \ldots
$$

(91)

At order $\alpha_s$ (remember that $\beta$ is of order $\alpha_s^2$) we get

$$
\frac{df_1}{d\mu^2} = 0 \Rightarrow f_1 = \text{constant} \equiv a_1.
$$

(92)

This is by itself a non-trivial result! It says that the evaluation of $R$ at one-loop is finite, all UV infinities must cancel without charge renormalization. If they didn’t cancel, $f_1$ would depend explicitly on $\mu$. As
we saw at the beginning, this is a consequence of the non-renormalization of the electric charge. At order \( \alpha_s^2 \) we have:

\[
\beta(\alpha_s) f_1(t) + \alpha_s^2 \frac{df_2}{d \log \mu^2} = 0 \Rightarrow f_2 = b_0 \log \frac{\mu^2}{s} + a_2 \text{ (integration constant)} .
\] (93)

So up to order \( \alpha_s^2 \) we have:

\[
R = 1 + a_1 \alpha_s + a_1 b_0 \alpha_s^2 \log \frac{\mu^2}{s} + a_2 \alpha_s^2 + \ldots
\] (94)

Notice that the requirement of renormalization group invariance allows us to know the coefficient of the logarithmic term at 2-loops without having to carry out the explicit 2-loop calculation! It is also important to notice that in the limit of high energy, \( s \to \infty \), the logarithmic term of the two-loop contribution becomes very large, and this piece becomes numerically of order \( \alpha_s \) as soon as \( \log \frac{s}{\mu^2} \gtrsim 1/b_0 \alpha_s \). You can easily check that renormalization scale invariance requires the presence of such logs at all orders of PT, in particular:

\[
f_{(n)}(t) = a_1 \left[ b_0 \log \frac{\mu^2}{s} \right]^n + \ldots
\] (95)

We can collect all these logs as follows:

\[
R = 1 + a_1 \alpha_s \left[ 1 + \alpha_s b_0 \log \frac{\mu^2}{s} + (\alpha_s b_0 \log \frac{\mu^2}{s})^2 + \ldots \right] + a_2 \alpha_s^2 + \ldots
\] (96)

\[
= 1 + a_1 \frac{\alpha_s(\mu)}{1 + \alpha_s(\mu) b_0 \log \frac{\mu^2}{s}} + a_2 \alpha_s^2 + \ldots \equiv 1 + a_1 \alpha_s(s) + a_2 \alpha_s^2 + \ldots
\] (97)

In fact:

\[
\frac{\alpha_s(\mu)}{1 + \alpha_s(\mu) b_0 \log \frac{\mu^2}{s}} = \frac{1}{b_0 \log \frac{\mu^2}{s} + \alpha_s(b_0 \log \frac{\mu^2}{s})} = \frac{1}{b_0 \log \frac{\mu^2}{s}} \equiv \alpha_s(s)
\] (98)

RG invariance constrains the form of higher-order corrections. All of the higher-order logarithmic terms are determined in terms of lower-order finite coefficients. They can be resummed by simply setting the scale of \( \alpha_s \) to \( s \). You can check by yourself that this will work also for the higher-order terms, such as those proportional to \( a_2 \). So the final result has the form:

\[
R = 1 + a_1 \alpha_s(s) + a_2 \alpha_s^2(s) + a_3 \alpha_s^3(s) + \ldots
\] (99)

Of course \( a_1, a_2, \ldots \) have to be determined by an explicit calculation. However, the truncation of the series at order \( n \) has now an accuracy which is truly of order \( \alpha_s^{n+1} \), contrary to before when higher-order terms were as large as lower-order ones. The explicit calculation has been carried out up to the \( a_3 \) coefficient, in particular,

\[
a_1 = \frac{3}{4} \frac{C_F}{\pi} \equiv \frac{1}{\pi}.
\] (100)

The formal proof of the previous equation can be obtained by showing that the general form of the equation

\[
\left[ \frac{\mu^2}{\partial \mu^2} + \beta(\alpha_s) \right] R(\alpha_s, \frac{s}{\mu^2}) = 0 ,
\] (101)

is given by

\[
\left\{ \begin{array}{l}
R(\alpha_s(s), 1) , \text{ with} \\
\frac{d \alpha_s}{d \log \frac{s}{\mu^2}} = \beta(\alpha_s) .
\end{array} \right.
\] (102)

58
QCD IN $e^+e^-$ COLLISIONS

$e^+e^-$ collisions provide one of the cleanest environments in which to study applications of QCD at high energy. This is the place where theoretical calculations have today reached their best accuracy, and where experimental data are the most precise, especially thanks to the huge statistics accumulated by LEP, LEP2 and SLC. The key process is the annihilation of the $e^+e^-$ pair into a virtual photon or $Z^0$ boson, which will subsequently decay to a $q\bar{q}$ pair. $e^+e^-$ collisions have therefore the big advantage of providing an almost point-like source of quark pairs, so that, contrary to the case of interactions involving hadrons in the initial state, we at least know very precisely the state of the quarks at the beginning of the interaction process.

Nevertheless, it is by no means obvious that this information is sufficient to predict the properties of the hadronic final state. We all know that this final state is clearly not simply a $q\bar{q}$ pair, but some high-multiplicity set of hadrons. It is therefore not obvious that a calculation done using the simple picture $e^+e^- \rightarrow q\bar{q}$ will have anything to do with reality. For example, one may wonder why don’t we need to calculate $\sigma(e^+e^- \rightarrow q\bar{q}g\ldots g\ldots)$ for all possible gluon multiplicities to get an accurate estimate of $\sigma(e^+e^- \rightarrow \text{hadrons})$. And since in any case the final state is not made of $q$’s and $g$’s, but of $\pi$’s, $K$’s, $\rho$’s, etc., why would $\sigma(e^+e^- \rightarrow q\bar{q}g\ldots g)$ be enough?

The solution to this puzzle lies both in a question of time and energy scales, and in the dynamics of QCD. When the $q\bar{q}$ pair is produced, the force binding $q$ and $\bar{q}$ is proportional to $\alpha_s(s)$ ($s$ being the $e^+e^-$ centre-of-mass energy). Therefore it is weak, and $q$ and $\bar{q}$ behave to good approximation like free particles. The radiation emitted in the first instants after the pair creation is also perturbative, and it will stay so until a time after creation of the order of $(1\ \text{GeV})^{-1}$, when radiation with wavelengths $\gtrsim (1\ \text{GeV})^{-1}$ starts being emitted. At this scale the coupling constant is large, non-perturbative phenomena and hadronization start playing a rôle. However, as we will show, colour emission during the perturbative evolution organizes itself in such a way as to form colour-neutral, low mass, parton clusters highly localized in phase-space. As a result, the complete colour-neutralization (i.e., the hadronization) does not involve long-range interactions between partons far away in phase-space. This is very important, because the forces acting among coloured objects at this time scale would be huge. If the perturbative evolution were to separate far apart colour-singlet $q\bar{q}$ pairs, the final-state interactions taking place during the hadronization phase would totally upset the structure of the final state. As an additional result of this “pre-confining” evolution, memory of where the local colour-neutral clusters came from is totally lost. So we expect the properties of hadronization to be universal: a model that describes hadronization at a given energy will work equally well at some other energy. Furthermore, so much time has passed since the original $q\bar{q}$ creation, that the hadronization phase cannot significantly affect the total hadron production rate. Perturbative corrections due to the emission of the first hard partons should be calculable in PT, providing a finite, meaningful cross-section.

The nature of non-perturbative corrections to this picture can be explored. One can prove for example that the leading correction to the total rate $R_{e^+e^-}$ is of order $F/s^2$, where $F \propto \langle 0 | \alpha_s F_{\mu\nu} \gamma^\mu | 0 \rangle$ is the so-called gluon condensate. Since $F \sim \mathcal{O}(1\ \text{GeV}^4)$, these NP corrections are usually very small. For example, they are of $\mathcal{O}(10^{-8})$ at the $Z^0$ peak! Corrections scaling like $\Lambda^2/s$ or $\Lambda/\sqrt{s}$ can nevertheless appear in other less inclusive quantities, such as event shapes or fragmentation functions.

We now come back to the perturbative evolution, and will devote the first part of this lecture to justifying the picture given above. In the second half we shall discuss jet cross-sections and shape variables.
4.1. Soft gluon emission

Emission of soft gluons plays a fundamental role in the evolution of the final state [6,15]. Soft gluons are emitted with large probability, since the emission spectrum behaves like \( dE/E \), typical of bremsstrahlung even in QED. They provide the seed for the bulk of the final-state multiplicity of hadrons. The study of soft-gluon emission is simplified by the simplicity of their couplings. Being soft (i.e., long wavelength) they are insensitive to the details of the very-short-distance dynamics: they cannot distinguish features of the interactions which take place on time scales shorter than their wavelength. They are also insensitive to the spin of the partons: the only feature they are sensitive to is the colour charge. To prove this let us consider soft-gluon emission in the \( q \bar{q} \) decay of an off-shell photon:

\[
\begin{align*}
A_{\text{soft}} &= \bar{u}(p)\epsilon(k) \frac{-i}{p + k} \Gamma^\mu v(\bar{p}) \lambda^a_{ij} + \bar{u}(p) \Gamma^\mu \frac{i}{p + k} (ig)\epsilon(k) v(\bar{p}) \lambda^a_{ij} \\
&= \left[ \frac{g}{2p \cdot k} \bar{u}(p)\epsilon(k) (\bar{p} + k)\Gamma^\mu v(\bar{p}) - \frac{g}{2p \cdot k} \bar{u}(p) \Gamma^\mu (\bar{p} + k)\epsilon(k) v(\bar{p}) \right] \lambda^a_{ij} .
\end{align*}
\]

I used the generic symbol \( \Gamma^\mu \) to describe the interaction vertex with the photon to stress the fact that the following manipulations are independent of the specific form of \( \Gamma^\mu \). In particular, \( \Gamma^\mu \) can represent an arbitrarily complicated vertex form factor. Neglecting the factors of \( k \) in the numerators (since \( k \ll p, \bar{p} \), by definition of soft) and using the Dirac equations, we get:

\[
A_{\text{soft}} = g\lambda^a_{ij} \left( \frac{p \cdot \epsilon}{p \cdot k} - \frac{\bar{p} \epsilon}{\bar{p} \cdot k} \right) A_{\text{Born}} .
\]

We then conclude that soft-gluon emission factorizes into the product of an emission factor, times the Born-level amplitude. From this exercise, one can extract general Feynman rules for soft-gluon emission:

\[
P \cdot j \quad p \cdot i = g \lambda^a_{ij} 2p^\mu
\]

**Exercise:** Derive the \( g \to gg \) soft-emission rules:

\[
P \cdot j \quad p \cdot i = ig f^{abc} 2p^\mu g^{\nu\rho}
\]

**Example:** Consider the “decay” of a virtual gluon into a quark pair. One more diagram should be added to those considered in the case of the electroweak decay. The fact that the quark pair is not in a colour-singlet state anymore makes things a bit more interesting:
\[ k_{\rightarrow 0} = \sum_{a,b,i,j} (\lambda^a \lambda^b)_{ij} \]

\[ = \sum_{a,b,i,j} \left[ (\lambda^a \lambda^b)_{ij}^2 - (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b,i,j} \left[ (\lambda^a \lambda^b)_{ij}^2 - (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b,i,j} \left[ (\lambda^a \lambda^b)_{ij}^2 - (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b,i,j} \left[ (\lambda^a \lambda^b)_{ij}^2 - (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b,i,j} \left[ (\lambda^a \lambda^b)_{ij}^2 - (\lambda^a \lambda^b)_{ij} \right] \]

The two factors correspond to the two possible ways colour can flow in this process:

\[ a \quad b \quad j \quad i \quad a \quad b \quad j \quad i \]

In the first case the antiquark (colour label \( j \)) is colour connected to the soft gluon (colour label \( b \)), and the quark (colour label \( i \)) is connected to the decaying gluon (colour label \( a \)). In the second case, the order is reversed. The two emission factors correspond to the emission of the soft gluon from the antiquark, and from the quark line, respectively. When squaring the total amplitude, and summing over initial and final-state colours, the interference between the two pieces is suppressed by \( 1/N^2 \) relative to the individual squares:

\[ \sum_{a,b,i,j} (\lambda^a \lambda^b)_{ij} = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

\[ = \sum_{a,b} \left[ (\lambda^a \lambda^b)_{ij} \right] \]

As a result, the emission of a soft gluon can be described, to the leading order in \( 1/N^2 \), as the incoherent sum of the emission form the two colour currents.

### 4.2. Angular ordering for soft-gluon emission

The results presented above have important consequences for the perturbative evolution of the quarks. A key property of the soft-gluon emission is the so-called angular ordering. This phenomenon consists in the continuous reduction of the opening angle at which successive soft gluons are emitted by the evolving quark. As a result, this radiation is confined within smaller and smaller cones around the quark direction, and the final state will look like a collimated jet of partons. In addition, the structure of the colour flow during the jet evolution forces the \( q\bar{q} \) pairs which are in a colour-singlet state to be close in phase-space, thereby achieving the pre-confinement of colour-singlet clusters alluded to at the beginning of the lecture.
Let us start first by proving the property of colour ordering. Consider the $q\bar{q}$ pair produced by the decay of a rapidly moving virtual photon. The amplitude for the emission of a soft gluon was given in Eq. (104). Squaring, summing over colours and including the gluon phase-space we get the following result:

$$
\frac{d\sigma}{d^3k} = \sum |A_{soft}|^2 \frac{d^3k}{(2\pi)^3 2k^0} \sum |A_0|^2 \frac{-2p^\mu p'^\nu}{(pk)(p'k)} \epsilon^*_\nu \epsilon_{\mu} \sum_{\alpha\beta} C_{\alpha\beta} \frac{d^3k}{(2\pi)^3 2k^0} \geq \frac{2(p\bar{p})}{(pk)(p'k)} g^2 C_f \frac{1}{8\pi^2} \frac{k^0 d\phi}{2\pi} \frac{d\theta}{8\pi^2} \frac{d^3k}{(2\pi)^3 2k^0}
$$

$$
= \frac{d\sigma_0}{(p\bar{p})/(p'k)} g^2 C_f \left(\frac{d\phi}{2\pi}\right) \frac{k^0 d\phi}{8\pi^2} \frac{d\theta}{8\pi^2} \frac{d^3k}{(2\pi)^3 2k^0}
$$

where $\theta_{\alpha\beta} = \theta_\alpha - \theta_\beta$, and $i, j, k$ refer to the $q, \bar{q}$ and gluon directions, respectively. We can write the following identity:

$$
\frac{1}{1 - \cos \theta_{ij}} \frac{1}{(1 - \cos \theta_{ik})(1 - \cos \theta_{jk})} = \frac{1}{2} \left[ \frac{\cos \theta_{jk} - \cos \theta_{ij}}{(1 - \cos \theta_{ik})(1 - \cos \theta_{jk})} + \frac{1}{1 - \cos \theta_{ik}} \right] + \frac{1}{2} [i \leftrightarrow j] \equiv W_{(i)} + W_{(j)} .
$$

We would like to interpret the two functions $W_{(i)}$ and $W_{(j)}$ as radiation probabilities from the quark and antiquark lines. Each of them is in fact only singular in the limit of gluon emission parallel to the respective quark:

$$
W_{(i)} \rightarrow \text{finite if } k \parallel j \ (\cos \theta_{jk} \rightarrow 1) ,
$$

$$
W_{(j)} \rightarrow \text{finite if } k \parallel i \ (\cos \theta_{ik} \rightarrow 1) .
$$

The interpretation as probabilities is however limited by the fact that neither $W_{(i)}$ nor $W_{(j)}$ are positive definite. However, you can easily prove that

$$
\int d\phi \frac{1}{2\pi} W_{(i)} = \left\{ \begin{array}{ll}
1 - \cos \theta_{ik} & \text{if } \theta_{ik} < \theta_{ij} , \\
0 & \text{otherwise} ,
\end{array} \right.
$$

where the integral is the azimuthal average around the $q$ direction. A similar result holds for $W_{(j)}$:

$$
\int d\phi \frac{1}{2\pi} W_{(j)} = \left\{ \begin{array}{ll}
1 - \cos \theta_{jk} & \text{if } \theta_{jk} < \theta_{ij} , \\
0 & \text{otherwise} .
\end{array} \right.
$$

As a result, the emission of soft gluons outside the two cones obtained by rotating the antiquark direction around the quark’s, and vice versa, averages to 0. Inside the two cones, one can consider the radiation from the emitters as being uncorrelated. In other words, the two colour lines defined by the quark and antiquark currents act as independent emitters, and the quantum coherence (i.e. the effects of interference between the two graphs contributing to the gluon-emission amplitude) is accounted for by constraining the emission to take place within those fixed cones.

If one repeats now the exercise for emission of one additional gluon, one will find the same angular constraint, but this time applied to the colour lines defined by the previously established antenna. As shown in the previous subsection, the $q\bar{q}g$ state can be decomposed at the leading order in $1/N$ into two independent emitters, one given by the colour line flowing from the gluon to the quark, the other given by the colour line flowing from the antiquark to the gluon. So the emission of the additional gluon will
be constrained to take place either within the cone formed by the quark and the gluon, or within the cone formed by the gluon and the antiquark. Either way, the emission angle will be smaller than the angle of the first gluon emission. This leads to the concept of angular ordering, with successive emission of soft gluons taking place within cones which get smaller and smaller.

The fact that colour always flows directly from the emitting parton to the emitted one, the collimation of the jet, and the softening of the radiation emitted at later stages, ensure that partons forming a colour-singlet cluster are close in phase-space. As a result, hadronization (the non-perturbative process that will bind together colour-singlet parton pairs) takes place locally inside the jet and is not a collective process: only pairs of nearby partons are involved. The inclusive properties of jets (e.g. the particle multiplicity, jet mass, jet broadening, etc.) are independent of the hadronization model, up to corrections of order \((\Lambda/\sqrt{s})^n\) (for some integer power \(n\), which depends on the observable), with \(\Lambda \lesssim 1\) GeV.

### 4.3. Jet rates

We now present explicit calculations of interesting observables. For simplicity, we will work with the soft-gluon approximation for the matrix elements and the phase-space. As a result, the correction to the differential \(e^+e^- \to q\bar{q}\) cross-section from one-gluon emission becomes:

\[
d\sigma_g = \sigma_0 \frac{2\alpha_s}{\pi} C_F \frac{d^2k}{k_0} \frac{d\cos \theta}{1 - \cos^2 \theta}, \quad \text{with } \sigma_0 = \text{Born amplitude.} \tag{118}
\]

In this equation we used the fact that in the soft-\(g\) limit the \(q\) and \(\bar{q}\) are back-to-back, and

\[
q \cdot \bar{q} = 2q_0 \bar{q}_0, \quad q \cdot k = q_0 k_0 (1 - \cos \theta), \quad \bar{q} k = \bar{q}_0 k_0 (1 + \cos \theta). \tag{119}
\]

Notice the presence in \(d\sigma_g\) of soft and collinear singularities. They will have to cancel in the total cross-section which, as we saw in the previous lecture, is finite. They do indeed cancel against the contribution to the total cross-section coming from the virtual correction diagram, where a gluon is exchanged between the two quarks. In the total cross-section (and for other sufficiently inclusive observables) the final states produced by the virtual diagrams and by the real emission diagrams in the soft or collinear limit are the same, and both contribute. In order for the total cross-section to be finite, the virtual contribution will need to take the following form:

\[
\frac{d^2\sigma_v}{dk_0 d\cos \theta} = -\sigma_0 \frac{2\alpha_s}{\pi} C_F \int_0^{\sqrt{s}/2} \frac{d^2k_0}{k_0} \int_{-1}^1 \frac{d\cos \theta'}{(1 - \cos^2 \theta')} \times \frac{1}{2} \delta(k_0) \left[ \delta(1 - \cos \theta) + \delta(1 + \cos \theta) \right] \tag{120}
\]

plus finite corrections. In this way:

\[
\int_0^{\sqrt{s}/2} \frac{d^2\sigma_g}{dk_0 d\cos \theta} + \frac{d^2\sigma_v}{dk_0 d\cos \theta} = \text{finite}. \tag{121}
\]

With the form of the virtual corrections available (at least in this simplified soft-gluon-dominated approximation) we can proceed and calculate other quantities.

Jets are usually defined as clusters of particles close-by in phase-space. A typical jet definition distributes particles in sets of invariant mass smaller than a given parameter \(M\), requiring that one particle only belongs to one jet, and that no other particles (or jets) can be added to a given jet without its mass exceeding \(M\). In the case of a three-particle final state, such as the one we are studying, we get three-jet events if \((q + k)^2, (\bar{q} + k)^2\) and \((q + \bar{q})^2\) are all larger than \(M^2\). We will have two-jet events when at least one of these quantities gets smaller than \(M^2\). For example emission of a gluon near the direction of the quark, with \(2qk = 2q_0 k_0 (1 - \cos \theta) < M^2\), defines a two-jet event, one jet being given by the \(\bar{q}\), the other by the system \(q + k\).
One usually introduces the parameter \( y = M^2/s \), and studies the jet multiplicity as a function of \( y \). Let us calculate the two- and three-jet rates at order \( \alpha_s \). The phase-space domain for two-jet events is given by two regions. The first one is defined by \( 2qk = 2q_0k_0(1 - \cos \theta) < ys \). This region consists of two parts:

\[
(I)_a : \quad \{ \begin{array}{l}
  k_0 < y\sqrt{s} \\
  0 < \cos \theta < 1
\end{array} \oplus \quad (I)_b : \quad \{ \begin{array}{l}
  k_0 > y\sqrt{s} \\
  1 - \frac{y\sqrt{s}}{k_0} < \cos \theta < 1
\end{array}, \tag{122}
\]

(I)_a corresponds to soft gluons at all angles smaller than \( \pi/2 \) (i.e. in the quark emisphere), while (I)_b corresponds to hard gluons emitted at small angles from the quark.

The second region, (II), is analogous to (I), but the angles are now referred to the direction of the antiquark. The integrals of \( d\sigma \) over (I) and (II) are of course the same. The \( \mathcal{O}(\alpha_s) \) contribution to the two-jet rate is therefore given by:

\[
\frac{\sigma_{2\text{-jet}}^{(\alpha_s)}}{\sigma_0} = \frac{1}{\sigma_0} \left[ 2 \int_{(I)_a} d\sigma_g + 2 \int_{(I)_b} d\sigma_g + \int_{\text{virtual}} d\sigma_v \right]
\]

\[
= \frac{4\alpha_s C_F}{\pi} \left[ \int_{0}^{y\sqrt{s}/k_0} \frac{d\sigma_g}{1 - \cos^2 \theta} + \int_{y\sqrt{s}/k_0}^{1} \frac{d\sigma_g}{1 - \cos^2 \theta} \right] + \frac{2\alpha_s C_F}{\pi} \left[ (-) \log \frac{k_0}{y\sqrt{s}} + (\text{finite for } y \to 0) \right] = \frac{-\alpha_s C_F}{\pi} \log^2 2y. \tag{123}
\]

Including the Born contribution, which always gives rise to two and only two jets, we finally have:

\[
\sigma_{2\text{-jet}} = \sigma_0 \left[ 1 - \frac{\alpha_s C_F}{\pi} \log^2 y + \ldots \right],
\]

\[
\sigma_{3\text{-jet}} = \sigma_0 \frac{\alpha_s C_F}{\pi} \log^2 y + \ldots
\]

If \( y \to 0 \), \( \sigma_{3\text{-jet}} \) becomes larger than \( \sigma_{2\text{-jet}} \). If \( y \) is sufficiently small, we can even get \( \sigma_{2\text{-jet}} < 0 \)!

This is a sign that higher-order corrections become important. In the soft-gluon limit, assuming that the emission of a second gluon will also factorize\(^2\), we can repeat the calculation at higher orders and obtain:

\[
\sigma_{2\text{-jet}} \simeq \sigma_0 \left[ 1 - \frac{\alpha_s C_F}{\pi} \log^2 y + \frac{1}{2!} \left( \frac{\alpha_s C_F}{\pi} \log^2 y \right)^2 + \ldots \right] = \sigma_0 e^{-\frac{\alpha_s C_F}{\pi} \log^2 y},
\]

\[
\sigma_{3\text{-jet}} \simeq \sigma_0 \frac{\alpha_s C_F}{\pi} \log^2 y \ e^{-\frac{\alpha_s C_F}{\pi} \log^2 y},
\]

\[
\vdots
\]

\(^2\)This is not true (see later on), but let us just accept it to see how things develop.
\[ \sigma_{(n+2)-jet} \sim \sigma_0 \frac{1}{n!} \left( \frac{\alpha_s C_F}{\pi} \log^2 y \right)^n e^{-\frac{\alpha_s C_F}{\pi} \log^2 y}. \] (124)

It is immediate to recognize in this series a Poisson distribution, leading to an average number of jets given by:

\[ \langle n_{jet} \rangle \simeq 2 + \frac{\alpha_s C_F}{\pi} \log^2 y. \] (125)

The smaller the resolution parameter \( y \), the smaller the mass of the jets, the larger the importance of higher order corrections. If we take the parameter \( M \) down to the scale of few hundred MeV hundred MeV (\( M \sim \Lambda_{QCD} \)), each particle gets identified with an independent jet. We can therefore estimate the \( s \)-dependence of the average multiplicity of particles produced:

\[ \langle n_{part} \rangle \sim \frac{C_F \alpha_s}{\pi} \log^2 \frac{s}{\Lambda^2} = \frac{C_F}{\pi b_0 \log \frac{\Lambda^2}{s}} \log^2 \frac{s}{\Lambda^2} \simeq \frac{C_F}{\pi b_0} \log \frac{s}{\Lambda^2}, \] (126)

The final state particle multiplicity grows with \( \log(s) \).

In practice, things are a bit more complicated than this. Once the first gluon is emitted, additional gluons can be emitted from it as well. Therefore the final-state multiplicity will be dominated by the emission of gluons from gluons. The analysis becomes more complicated (see e.g. Refs. [6,8] for the details), and the final result is:

\[ \langle n_{part}(s) \rangle \sim \exp \sqrt{\frac{2C_A}{\pi b_0} \log \frac{s}{\Lambda^2}} \] (127)

for the particle multiplicity, and

\[ \langle n_{jet}(y) \rangle = 2 + 2 \frac{C_F}{C_A} (\cosh \sqrt{\frac{\alpha_s C_A}{2\pi} \log^2 \frac{1}{y}} - 1) \sim \frac{C_F}{C_A} \exp \sqrt{\frac{\alpha_s C_A}{2\pi} \log^2 \frac{1}{y}} \] (128)

for the average jet multiplicity.

Other interesting quantities that can be calculated using the simple formulas we developed so far are the average jet mass and the thrust. To define the jet mass we just divide the final state into two emispheres, separated by the plane orthogonal to the thrust axis. We now call jets the two sets of particles on either side of the plane. The \( \langle m^2 \rangle \) of the jet is then given by

\[ \langle m^2_{jet} \rangle = \frac{1}{2\sigma_0} \left\{ \int_{(I)} (q+k)^2 d\sigma_y + \int_{(II)} (q+k)^2 d\sigma_y \right\}. \] (129)

The virtual correction does not enter here, since the pure \( q\bar{q} \) final state has jet masses equal to 0. The result of this simple computation leads to

\[ \langle m^2_{jet} \rangle = \frac{\alpha_s C_F}{\pi} \frac{s}{8}. \] (130)

Another interesting variable often used in experimental studies is the thrust \( T \), defined by:

\[ T = \max_{\hat{T}} \sum_i |\vec{p}_i \cdot \hat{T}| \bigg/ \sum_i |\vec{p}_i|, \]

where \( \hat{T} \) is the thrust axis, defined so as to maximize \( T \). For three-body final states, \( \hat{T} \) is the direction of
the highest-energy parton, and $T$ is proportional to twice its energy:

$$T = 2 \frac{\bar{q}_0}{\sqrt{s}} = 1 - \frac{(q + k)^2}{s} = 1 - \frac{m_{jet}^2}{s}. \quad (131)$$

As a result:

$$\langle 1 - T \rangle = \frac{\alpha_s C_F}{\pi}. \quad (132)$$

At LEP, $\langle 1 - T \rangle \simeq 0.12 \times \frac{4}{3} \simeq 0.05$. The terms neglected in the soft-gluon approximation we used throughout can be calculated, and give some small correction to the above results. Corrections will likewise come from higher-order effects. State-of-the-art calculations exist which evaluate all these “shape variables” (and more!) up to $O(\alpha_s^2)$ accuracy, including a full next-to-leading-log accurate resummation of higher-order logarithms (such as the log $1/y$ terms we encountered in the discussion of jet rates, or terms of the form $\log^n(1 - T)$ which appear at higher orders in the evaluation of the thrust distributions). These calculations allow a reliable estimate of several different observables directly proportional to $s$, and provide the theoretical input for the extraction of $\alpha_s$ from the LEP QCD data [8].

Notice that non-perturbative corrections proportional to $\frac{N_C}{\sqrt{s}}$, with $N_C \sim 1$ GeV, can have a significant impact on the extraction of $\alpha_s$. For example, a $\frac{N_C}{\sqrt{s}}$ correction to $\langle 1 - T \rangle$ would be a 20% effect:

$$\frac{\Lambda}{\sqrt{s}} \sim 0.01, \quad \langle 1 - T \rangle_{PT} \simeq 0.05.$$

Indeed one measures $\langle 1 - T \rangle_{LEP} = 0.068 \pm 0.003$, vs the full PT QCD prediction of 0.055 (using $\alpha_s = 0.120$).

5. **QCD AND THE PROTON STRUCTURE AT LARGE $Q^2$**

The understanding of the structure of the proton at short distances is one of the key ingredients to be able to predict cross-section for processes involving hadrons in the initial state. All processes in hadronic collisions, even those intrinsically of electroweak nature such as the production of $W/Z$ bosons or photons, are in fact induced by the quarks and gluons contained inside the hadron. In this lecture I will introduce some important concepts, such as the notion of partonic densities of the proton, and of parton evolution. These are the essential tools used by theorists to predict production rates for hadronic reactions.

The idea that the parton language [1] and the use of perturbative QCD can be used to describe the structure of the proton at short distances was developed in the late 60’s and early 70’s (for a nice review, see Ref. [17]). While I will not provide you with a rigorous proof of the legitimacy of this approach, I will try to justify it qualitatively to make it sound at least plausible. I will then proceed to extract some results based on the application of perturbative QCD to lepton-hadron interactions.

5.1. **The parton model**

We all know that quarks are deeply bound inside the proton. It is important to realise, however, that the binding forces responsible for the quark confinement are due to the exchange of rather soft gluons. If a quark were to exchange a hard virtual gluon with another quark, in fact, the recoil would tend to break the proton apart. It is easy to verify that the exchange of gluons with virtuality larger than $Q$ is then proportional to some large power of $m_p/Q$, $m_p$ being the proton mass. Since the gluon coupling
constant gets smaller at large $Q$, exchange of hard gluons is significantly suppressed. As a result, the typical time scale for quarks inside the proton to interact among themselves is of the order of $1/m_p$, or longer. If we probe the proton with an off-shell photon, the interaction should take place during the limited lifetime of the virtual photon, given by the inverse of its virtuality as a result of the Heisenberg principle. Once the photon gets “inside” the proton and meets a quark, the struck quark has no time to negotiate a coherent response with the other quarks, because the time scale for it to “talk” to its pals is too long compared with the duration of the interaction with the photon itself. As a result, the struck quark has no option but to interact with the photon as if it were a free particle.

The one thing that the above picture does not tell us, obviously, is in which precise state the quark was once it got struck by the photon. This depends on the internal wave function of the proton, which perturbative QCD cannot easily predict. We can however say that the wave function of the proton, and therefore the state of the “free” quark, are determined by the dynamics of the soft-gluon exchanges inside the proton itself. Since the time scale of this dynamics is long relative to the time scale of the photon-quark interaction, we can safely argue that the photon sees to good approximation a static snapshot of the proton’s inner guts. In other words, the state of the quark had been prepared long before the photon arrived. This also suggests that the state of the quark will not depend on the precise nature of the external probe, provided the time scale of the hard interaction is very short compared to the time it would take for the quark to readjust itself. As a result, if we could perform some measurement of the quark state using, say, a virtual-photon probe, we could then use this knowledge on the state of the quark to perform predictions for the interaction of the proton with any other probe (e.g. a virtual $W$ or even a gluon from an opposite beam of hadrons).

In order to make the measurement of the proton structure as simple as possible, it is therefore wise to use a probe as simple as possible. A virtual photon emitted from a beam of high-energy electrons provides such a probe. The relative process is called deeply inelastic scattering (DIS), and was historically the first phenomenon which led people to introduce the concept of partons [2].

Assuming the parton picture outlined above, we can describe the cross-section for the interaction of the virtual photon with the proton as follows:

$$\sigma_0 = \int_0^1 dx \sum_i e_i^2 f_i(x) \hat{\sigma}_0(\gamma^* q_i \rightarrow q'_i, x),$$

where the 0 subscript anticipates that this description represents a leading order approximation. In the above equation, $f_i(x)$ represents the density of quarks of flavour $i$ carrying a fraction $x$ of the proton momentum. The hatted cross-section represents the interaction between the photon and a free (massless) quark:

$$\hat{\sigma}_0(\gamma^* q_i \rightarrow q'_i) = \frac{1}{f_{\text{uxx}}} \sum |M_0(\gamma^* q \rightarrow q')|^2 \frac{d^3p'}{(2\pi)^3 2p'_0} (2\pi)^4 \delta^4(p' - q - p)$$

$$= \frac{1}{f_{\text{uxx}}} \sum |M_0|^2 2\pi \delta(p'^2).$$

Using $p' = xP + q$, where $P$ is the proton momentum, we get

$$(p')^2 = 2xP \cdot q + q^2 \equiv 2xP \cdot q - Q^2,$$

$$\hat{\sigma}_0(\gamma^* q \rightarrow q') = \frac{2\pi}{f_{\text{uxx}}} \sum |M_0|^2 \frac{1}{2P \cdot q} \delta(x - x_{bj}),$$

The fact that the coupling decreases at large $Q$ plays a fundamental role in this argument. Were this not true, the parton picture could not be used!
where \( x_{bj} = \frac{Q^2}{2 P \cdot q} \) is the so-called Bjorken-\( x \) variable. Finally:

\[
\sigma_0 = \frac{2 \pi}{f_l u x} \sum_i \frac{|M_i|^2}{Q^2} \sum_j x_{bj} f_i(x_{bj}) e_i^2 = \frac{2 \pi}{f_l u x} \sum |M_0|^2 \frac{Q^2}{Q^2} F_2(x_{bj}) .
\] (137)

The measurement of the inclusive \( ep \) cross-section as a function of \( Q^2 \) and \( P \cdot q \) (= \( m_p(E' - E) \) in the proton rest frame, with \( E' = \) energy of final-state lepton and \( E = \) energy of initial-state lepton) probes the quark momentum distribution inside the proton.

### 5.2. Parton evolution

Let us now study the QCD corrections to the LO parton-model description of DIS. This study will exhibit many important aspects of QCD (structure of collinear singularities, renormalization-group invariance) and will take us to an important element of the DIS phenomenology, namely scaling violations. We start from real-emission corrections to the Born level process:

\[
\begin{array}{c}
\text{\( q \)} \\
\text{\( p \)} \\
\text{\( k \)}
\end{array}
+ \begin{array}{c}
\text{\( q \)} \\
\text{\( p' \)} \\
\text{\( k \)}
\end{array}
\] (138)

The first diagram is proportional to \( 1/(p-k)^2 = 1/2(pk) \), which diverges when \( k \) is emitted parallel to \( p \):

\[
p \cdot k = p^0 k^0 \left(1 - \cos \theta\right) \cos \theta \rightarrow 0 .
\] (139)

The second diagram is also divergent, if \( k \) is emitted parallel to \( p' \). This second divergence turns out to be harmless, since we are summing over all possible final states. Whether the final-state quark keeps all of its energy, or whether it decides to share it with a gluon emitted collinearly, an inclusive final-state measurement will not care. The collinear divergence can then be cancelled by a similar divergence appearing in the final-state quark self-energy corrections.

The first divergence is more serious, since from the point of view of the incoming photon (which only sees the quark, not the gluon) it does make a difference whether the momentum is all carried by the quark or is shared between the quark and the gluon. This means that no cancellation between collinear singularities in the real emission and virtual emission is possible. So let us go ahead, calculate explicitly the contribution of these diagrams, and learn how to deal with their singularities.

First of all note that while the second diagram is not singular in the region \( k \cdot p \rightarrow 0 \), its interference with the first one is. It is possible, however, to select a gauge for which the interference of the two diagrams is finite in this limit. You can show that the right choice is

\[
\sum \epsilon_\mu \epsilon^*_\nu(k) = -g_{\mu \nu} + \frac{k^\mu p'^\nu + k^\nu p'^\mu}{k \cdot p'} .
\] (140)

Notice that in this gauge not only \( k \cdot \epsilon(k) = 0 \), but also \( p' \cdot \epsilon(k) = 0 \). The key to getting to the end of a QCD calculation in a finite amount of time is choosing a proper gauge (which we just did) and the proper parametrization of the momenta involved. In our case, since we are interested in isolating the
region where $k$ becomes parallel with $p$, it is useful to set
\[ k_\mu = (1 - z)p_\mu + \beta p'_\mu + (k_\perp)_\mu , \] (141)
with $k_\perp \cdot p = k_\perp \cdot p' = 0$. $\beta$ is obtained by imposing
\[ k^2 = 0 = 2\beta(1 - z)p \cdot p' + k^2_\perp . \] (142)
Defining $k^2_\perp = -k^2_t$, we then get
\[ \beta = \frac{k^2_t}{2(\epsilon p' / (1 - z) \),} \] (143)
\[ k_\mu = (1 - x)p_\mu + \frac{k^2_t}{2(1 - x)p \cdot p'} p'_\mu + (k_\perp)_\mu , \] (144)
$(k_\perp)_\mu$ is therefore the gluon momentum vector transverse to the incoming quark, in a frame where $\gamma^x$ and $q$ are aligned. $k_t$ is the value of this transverse momentum. We also get
\[ k \cdot p = \beta p \cdot p' = \frac{k^2_t}{2(1 - z)} \text{ and } k \cdot p' = (1 - z)p \cdot p' . \] (145)
As a result $(p - k)^2 = -k^2_t / (1 - z)$. The amplitude for the only diagram carrying the initial-state singularity is:
\[ M_g = ig\lambda^\alpha_\beta \bar{u}(p')\Gamma \frac{\hat{p} - \hat{k}}{(p - k)^2} \hat{e}(k)u(p) , \] (146)
(where we introduced the notation $\hat{a} \equiv \hat{b} \equiv a_\mu \gamma^\mu$). We indicated by $\Gamma$ the interaction vertex with the external current $q$. It is important to keep $\Gamma$ arbitrary, because we would like to get results which do not depend on the details of the interaction with the external probe. It is important that the singular part of the QCD correction, and therefore its renormalization, be process independent. Only in this way we can hope to achieve a true universality of the parton densities! So we will keep $\Gamma$ generic, and make sure that our algebra does not depend on its form, at least in the $p \cdot k \rightarrow 0$ limit. Squaring the most singular part of the amplitude, and summing over colours and spins, we get:
\[ \sum_{g \text{ polariz. and colours}} |M_g|^2 = g^2 \sum_{a} \text{tr} (\lambda^a \lambda^a) \times \frac{1}{t^2} \times \sum_{\epsilon} \text{Tr} [p' \Gamma (\hat{p} - \hat{k}) \hat{e} p \hat{e}^\ast (\hat{p} - \hat{k}) \Gamma^+ ] \times \sum_{\epsilon} \text{Tr} [p' \Gamma (\hat{p} - \hat{k}) \hat{e} p \hat{e}^\ast (\hat{p} - \hat{k}) \Gamma^+ ] , \] (147)
with $t = (p - k)^2 = -k^2_t / (1 - z)$. Let us look first at
\[ \sum_{\epsilon} \hat{e} \hat{p} \hat{e}^\ast = \sum_{\epsilon} \epsilon_\mu \epsilon_\nu^* \gamma^\mu \hat{p} \gamma^\nu = -\gamma^\mu \hat{p} \gamma^\mu + \frac{1}{k \cdot p'} (\hat{p}' \hat{k} + \hat{k} \hat{p}') = \frac{2}{1 - z} (k + \beta \hat{p}') , \] (148)
(we used: $\hat{a} \hat{b} \hat{c} + \hat{c} \hat{b} \hat{a} = 2(a \cdot b) \hat{c} - 2(a \cdot c) \hat{b} + 2(b \cdot c) \hat{a}$ and some of the kinematical relations from the previous page). Then take
\[ (\hat{p} - \hat{k}) (k + \beta \hat{p}') (\hat{p} - \hat{k}) = (\hat{p} - \hat{k}) k (\hat{p} - \hat{k}) + \beta (\hat{p} - \hat{k}) \hat{p}' (\hat{p} - \hat{k}) . \] (149)
In the second term, proportional to $\beta$, we can approximate $\hat{k} = (1 - z)\hat{p}$. This is because the other pieces $(\beta \hat{p}' + \hat{k}_\perp)$ multiplied by $\beta$ would cancel entirely the $\frac{1}{t^2}$ singularity, and would only contribute a non-singular term, which we are currently neglecting. So Eq. (149) becomes
\[ \hat{p} \hat{k} \hat{p} + \beta z^2 \hat{p} \hat{p} \hat{p} = 2(p \cdot k) \hat{p} + \beta z^2 2(p \cdot p') \hat{p} = 2(p \cdot k) (1 + z^2) \hat{p} \] (150)
and
\[ \sum |M_g|^2 = 2g^2 C_F \frac{(1-z)}{k_f^2} \left( \frac{1+z^2}{1-z} \right) N \text{Tr}[\not{p}\Gamma\not{p}\Gamma] . \] (151)

The last factor with the trace corresponds to the Born amplitude squared. So the one-gluon emission process factorizes in the collinear limit into the Born process times a factor which is independent of the beam’s nature! If we add the gluon phase-space:
\[ [dk] \equiv \frac{d^3k}{(2\pi)^32k^0} = \frac{dk_\parallel}{k^0} \frac{d\phi}{2\pi} \frac{dk^2}{8\pi^2} = \frac{dz}{(1-z)} \frac{1}{16\pi^2} dk^2_\perp , \] (152)
we get:
\[ \sum |M_g|^2 [dk] = \frac{dk^2}{k^2} dz \left( \frac{\alpha_s}{2\pi} \right) P_{qq}(z) \sum |M_0|^2 , \] (153)
where
\[ P_{qq}(z) = C_F \frac{1+z^2}{1-z} \] (154)
is the so-called Altarelli-Parisi splitting function for the $q \rightarrow q$ transition ($z$ is the momentum fraction of the original quark taken away by the quark after gluon emission). We are now ready to calculate the corrections to the parton-model cross-section:
\[ \sigma_g = \int dx f(x) \int \frac{dk^2}{k^2} \frac{dz}{z} \left( \frac{\alpha_s}{2\pi} \right) P_{qq}(z) \sum |M_0|^2 2\pi \delta(p'^2) . \] (155)

Using $(p')^2 = (p-k+q)^2 \sim (zp+q)^2 = (xzP+q)^2$ and
\[ \delta(p'^2) = \frac{1}{2P \cdot q} \frac{1}{z} \delta(x-x_b) = \frac{x_b}{z} \delta(x-x_b) , \] (156)
we finally obtain:
\[ \sigma_g = \frac{2\pi}{f_{lux}} \frac{\sum |M_0|^2}{Q^2} \sum \epsilon^2_i x_{bj} \frac{\alpha_s}{2\pi} \int \frac{dk^2}{k^2} \int \frac{dz}{z} P_{qq}(z) f_i \left( \frac{x_b}{z} \right) . \] (157)

We then find that the inclusion of the $O(\alpha_s)$ correction is equivalent to a contribution to the parton density:
\[ f_i(x) \rightarrow f_i(x) + \frac{\alpha_s}{2\pi} \int \frac{dk^2}{k^2} \int_0^1 \frac{dz}{z} P_{qq}(z) f_i \left( \frac{x}{z} \right) . \] (158)

Notice the presence of the integral $\int \frac{dk^2}{k^2}$. The upper limit of integration is proportional to $Q^2$. The lower limit is 0. Had we included a quark mass, the propagator would have behaved like $1/(k^2 + m^2)$. But the quark is bound inside the hadron, so we do not quite know what $m$ should be. Let us then assume that we cutoff the integral at a $k_\perp$ value equal to some scale $\mu_0$, and see what happens. The effective parton density becomes:
\[ f(x, Q^2) = f(x) + \log \left( \frac{Q^2}{\mu_0^2} \right) \frac{\alpha_s}{2\pi} \int \frac{dz}{z} P_{qq}(z) f \left( \frac{x}{z} \right) . \] (159)

The dependence on the scale $\mu_0$, which is a non-perturbative scale, can be removed by defining $f(x, Q^2)$...
in terms of the parton density \( f \) measured at a large, perturbative scale \( \mu^2 \):

\[
f(x, \mu^2) = f(x) + \log \left( \frac{\mu^2}{\mu_0^2} \right) \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{qq}(z) f \left( \frac{x}{z} \right).
\] (160)

We can then perform a subtraction, and write:

\[
f(x, Q^2) = f(x, \mu^2) + \log \left( \frac{Q^2}{\mu^2} \right) \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{qq}(z) f \left( \frac{x}{z} \right).
\] (161)

The scale \( \mu \) plays here a similar role to the renormalization scale introduced in the second lecture. Its choice is arbitrary, and \( f(x, Q^2) \) should not depend on it. Requiring this independence, we get the following “renormalization-group invariance” condition:

\[
\frac{df(x, Q^2)}{d \ln \mu^2} = \mu^2 \frac{df(x, \mu^2)}{d \mu^2} - \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{qq}(z) f \left( \frac{x}{z} \right) \equiv 0
\] (162)

and then

\[
\mu^2 \frac{df(x, \mu^2)}{d \mu^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{qq}(z) f \left( \frac{x}{z}, \mu^2 \right).
\] (163)

This equation is usually called the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equation. As in the case of the resummation of leading logarithms in \( R_{e^+e^-} \) induced by the RG invariance constraints, the DGLAP equation – which is the result of RG-invariance – resums a full tower of leading logarithms of \( Q^2 \).

**Proof:** Let us define \( t = \log \frac{Q^2}{\mu^2} \). We can then expand \( f(x, t) \) in powers of \( t \):

\[
f(x, t) = f(x, 0) + t \frac{df}{dt}(x, 0) + \frac{t^2}{2!} \frac{d^2 f}{dt^2}(x, 0) + \ldots
\] (164)

The first derivative is given by the DGLAP equation itself. Higher derivatives can be obtained by differentiating it:

\[
f''(x, t) = \frac{\alpha_s}{2\pi} \int \frac{dz}{z} P_{qq}(z) \frac{df}{dt} \left( \frac{x}{z}, t \right),
\]

\[
= \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} P_{qq}(z) \frac{\alpha_s}{2\pi} \int \frac{dz'}{z'} P_{qq}(z') f \left( \frac{x}{zz'}, t \right),
\]

\[
\vdots
\]

\[
f^{(n)}(x, t) = \frac{\alpha_s}{2\pi} \int_x^1 \ldots \frac{\alpha_s}{2\pi} \int_x^{zz'\ldots} \frac{dz(n)}{z(n)} P_{qq}(z(n)) f \left( \frac{x}{zz'\ldots}, t \right).
\] (165)

The \( n \)-th term in this expansion, proportional to \( (\alpha_s t)^n \), corresponds to the emission of \( n \) gluons (it is just the \( n \)-fold iteration of what we did studying the one-gluon emission case).

With similar calculations one can include the effect of the other \( O(\alpha_s) \) correction, originating from the splitting into a \( q\bar{q} \) pair of a gluon contained in the proton. With the addition of this term, the evolution equation for the density of the \( i \)th quark flavour becomes:

\[
\frac{df_i(x, t)}{dt} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left[ P_{qq}(z) f_i \left( \frac{x}{z}, t \right) + P_{qg}(z) f_g \left( \frac{x}{z}, t \right) \right], \quad \text{with } P_{qg} = \frac{1}{2} \left[ z^2 + (1 - z)^2 \right].
\] (166)

In the case of interactions with a coloured probe (say a gluon) we meet the following corrections, which
affect the evolution of the gluon density $f_g(x)$:

$$
\frac{df_g(x,t)}{dt} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dz}{z} \left[ P_{gq}(z) \sum_{i=q,\bar{q}} f_i \left( \frac{x}{z},t \right) + P_{gg}(z) f_g \left( \frac{x}{z},t \right) \right]
$$

with

$$P_{gq}(z) = P_{qq}(1-z) = C_F \frac{1 + (1-z)^2}{z} \quad \text{and} \quad P_{gg}(z) = 2C_A \left[ \frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right].$$

Defining the moments of an arbitrary function $g(x)$ as follows:

$$g_n = \int_0^1 dx \frac{1}{x^n} g(x),
$$

it is easy to prove that the evolution equations turn into ordinary linear differential equations:

$$\frac{df_i^{(n)}}{dt} = \frac{\alpha_s}{2\pi} [P_{qq}^{(n)} f_i^{(n)} + P_{gq}^{(n)} f_g^{(n)}], \quad (169)
$$

$$\frac{df_q^{(n)}}{dt} = \frac{\alpha_s}{2\pi} [P_{qq}^{(n)} f_q^{(n)} + P_{gq}^{(n)} f_g^{(n)}]. \quad (170)
$$

### 5.3. Properties of the evolution equations

We now study some general properties of these equations. It is convenient to introduce the concepts of valence ($V(x,t)$) and singlet ($\Sigma(x,t)$) densities:

$$V(x) = \sum_i f_i(x) - \sum_i \bar{f}_i(x),
$$

$$\Sigma(x) = \sum_i f_i(x) + \sum_i \bar{f}_i(x),
$$

where the index $\bar{i}$ refers to the antiquark flavours. The evolution equations then become:

$$\frac{dV^{(n)}}{dt} = \frac{\alpha_s}{2\pi} P_{qq}^{(n)} V^{(n)},
$$

$$\frac{d\Sigma^{(n)}}{dt} = \frac{\alpha_s}{2\pi} \left[ P_{qq}^{(n)} \Sigma^{(n)} + 2n_f P_{gq}^{(n)} f_g^{(n)} \right],
$$

$$\frac{df_g^{(n)}}{dt} = \frac{\alpha_s}{2\pi} \left[ P_{gq}^{(n)} \Sigma^{(n)} + P_{gg}^{(n)} f_g^{(n)} \right].
$$

Note that the equation for the valence density decouples from the evolution of the gluon and singlet densities, which are coupled among themselves. This is physically very reasonable, since in perturbation theory the contribution to the quark and the antiquark densities coming from the evolution of gluons (via their splitting into $q\bar{q}$ pairs) is the same, and will cancel out in the definition of the valence. The valence therefore only evolves because of gluon emission. On the contrary, gluons and $q\bar{q}$ pairs in the proton sea evolve into one another.

The first moment of $V(x)$, $V^{(1)} = \int_0^1 dx V(x)$, counts the number of valence quarks. We therefore expect it to be independent of $Q^2$:

$$\frac{dV^{(1)}}{dt} \equiv 0 = \frac{\alpha_s}{2\pi} P_{qq}^{(1)} V^{(1)} = 0.
$$

(176)
Since $V^{(1)}$ itself is different from 0, we obtain a constraint on the first moment of the splitting function: $P^{(1)}_{qq} = 0$. This constraint is satisfied by including the effect of the virtual corrections, which generate a contribution to $P_{qq}(z)$ proportional to $\delta(1 - z)$. This correction is incorporated in $P_{qq}(z)$ via the redefinition:

$$P_{qq}(z) \rightarrow \left(\frac{1 + z^2}{1 - z}\right) + \frac{1 + z^2}{1 - z} - \delta(1 - z) \int_0^1 dy \left(\frac{1 + y^2}{1 - y}\right) ,$$

(177)

where the + sign turns $P_{qq}(z)$ into a distribution. In this way, $\int_0^1 dz P_{qq}(z) = 0$ and the valence sum-rule is obeyed at all $Q^2$.

Another sum rule which does not depend on $Q^2$ is the momentum sum rule, which imposes the constraint that all of the momentum of the proton is carried by its constituents (valence plus sea plus gluons):

$$\int_0^1 dx x \left[ \sum_{i,s} f_i(x) + f_g(x) \right] \equiv \Sigma^{(2)} + f_g^{(2)} = 1 .$$

(178)

Once more this relation should hold for all $Q^2$ values, and you can prove by using the evolution equations that this implies:

$$P^{(2)}_{qq} + P^{(2)}_{gq} = 0 ,$$

(179)

$$P^{(2)}_{gg} + 2n_f P^{(2)}_{qg} = 0 .$$

(180)

You can check using the definition of second moment, and the explicit expressions of the $P_{qq}$ and $P_{gg}$ splitting functions, that the first condition is automatically satisfied. The second condition is satisfied by including the virtual effects in the gluon propagator, which contribute a term proportional to $(1 - z)$. It is a simple exercise to verify that the final form of the $P_{gg}(z)$ splitting function, satisfying Eq. (180), is:

$$P_{gg} \rightarrow 2CA \left\{ \frac{x}{(1 - x)_+} + \frac{1 - x}{x} + x(1 - x) \right\} + \delta(1 - x) \left[ \frac{11CA - 2n_f}{6} \right] .$$

(181)

5.4. Solution of the evolution equations

The evolution equations formulated in the previous section can be solved analytically in moment space. The boundary conditions are given by the moments of the parton densities at a given scale $\mu$, where in principle they can be obtained from a direct measurement. The solution at different values of the scale $Q$ can then be obtained by inverting numerically the expression for the moments back to $x$ space. The resulting evolved densities can then be used to calculate cross sections for an arbitrary process involving hadrons, at an arbitrary scale $Q$. We shall limit ourselves here to studying some properties of the analytic solutions, and will present and comment some plots obtained from numerical studies available in the literature.

As an exercise, you can show that the solution of the evolution equation for the valence density is the following:

$$V^{(n)}(Q^2) = V^{(n)}(\mu^2) \left[ \log \frac{Q^2}{\Lambda^2} \right]^{P^{(n)}_{qq}/2\pi b_0} = V^{(n)}(\mu^2) \left[ \frac{\alpha_s(\mu^2)}{\alpha_s(Q^2)} \right]^{P^{(n)}_{qq}/2\pi b_0} ,$$

(182)

where the running of $\alpha_s(\mu^2)$ has to be taken into account to get the right result. Since all moments $P^{(n)}$ are negative, the evolution to larger values of $Q$ makes the valence distribution softer and softer. This is physically reasonable, since the only thing that the valence quarks can do is to loose energy because of gluon emission.
The solutions for the gluon and singlet distributions \( f_g \) and \( \Sigma \) can be obtained by diagonalizing the \( 2 \times 2 \) system in Eqs. (174) and (175). We study the case of the second moments, which correspond to the momentum fractions carried by quarks and gluons separately. In the asymptotic limit \( \Sigma^{(2)} \) goes to a constant, and \( \frac{d\Sigma^{(2)}}{dt} = 0 \). Then, using the momentum sum rule:

\[
\begin{align*}
P_{qq}^{(2)} \Sigma^{(2)} + 2n_f P_{qg}^{(2)} f_g^{(2)} &= 0 , \\
\Sigma^{(2)} + f_g^{(2)} &= 1 .
\end{align*}
\]

The solution of this system is:

\[
\begin{align*}
\Sigma^{(2)} &= \frac{1}{1 + \frac{4C_F}{n_f}} \quad (= 15/31 \text{ for } n_f = 5) , \\
f_g^{(2)} &= \frac{4C_F}{4C_F + n_f} \quad (= 16/31 \text{ for } n_f = 5) .
\end{align*}
\]

As a result, the fraction of momentum carried by gluons is asymptotically approximately 50% of the total proton momentum. It is interesting to note that, experimentally, this asymptotic value is actually reached already at rather low values of \( Q^2 \). It was indeed observed already since the early days of the DIS experiments that only approximately 50% of the proton momentum was carried by charged constituents. This was one of the early evidences for the existence of gluons.

As I mentioned earlier, a complete solution for the evolved parton densities in \( x \) space can only be obtained from a numerical analysis. This work has been done in the past by several groups (see e.g. the discussions in Ref. [8]), and is continuously being updated by including the most up-to-date experimental results used for the determination of the input densities at a fixed scale. Figure 2a describes the up-quark valence momentum density at different scales \( Q \). Note the anticipated softening at large scales, and the clear \( \log Q^2 \) evolution. The most likely momentum fraction carried by a valence up quark in the proton goes from \( x \sim 20\% \) at \( Q = 3 \text{ GeV} \), to \( x \lesssim 10\% \) at \( Q = 1000 \text{ GeV} \). Notice finally that the density vanishes at small \( x \).

Figure 2b shows instead the gluon momentum density at different scales \( Q \). This time the density grows at small-\( x \), with an approximate \( g(x) \sim 1/x^{1+\delta} \) behaviour, and \( \delta > 0 \) slowly increasing at large \( Q^2 \). This low-\( x \) growth is due to the \( 1/x \) emission probability for the radiation of gluons, which was discussed in the previous lecture and which is represented by the \( 1/x \) factors in the \( P_{qg}(x) \) and
Fig. 3: Left: Sea up-quark momentum-density distribution, for different scales $Q$. Right: Momentum-density distribution for several parton species, at $Q = 1000$ GeV.

$P_{gg}(x)$ splitting functions. Figure 3a shows the up-quark sea momentum density at different scales $Q$. Shape and evolution match those of the gluon density, a consequence of the fact that sea quarks come from the splitting of gluons. Since the gluon-splitting probability is proportional to $\alpha_s$, the approximate ratio $\text{sea}/\text{gluon} \sim 0.1$ which can be obtained by comparing Figs. 2b and 3a is perfectly justified.

Finally, the momentum densities for gluons, up-sea, charm and up-valence distributions are shown in Fig. 3b for $Q = 1000$ GeV. Note here that $u_{\text{sea}}$ and charm are approximately the same at very large $Q$ and small $x$. The proton momentum is mostly carried by valence quarks and by gluons. The contribution of sea quarks is negligible.

6. QCD IN HADRONIC COLLISIONS

In hadronic collisions, all phenomena are QCD-related. The dynamics is more complex than in $e^+e^-$ or DIS, since both beam and target have a non-trivial partonic structure. As a result, calculations (and experimental analyses) are more complicated. QCD phenomenology is however much richer, and the higher energies available in hadronic collisions allow to probe the structure of the proton and of its constituents at the smallest scales attainable in a laboratory.

Contrary to the case of $e^+e^-$ and lepton-hadron collisions, where calculations are routinely available up to next-to-next-to-leading order (NNLO) accuracy, theoretical calculations for hadronic collisions are available at best with next-to-leading-order (NLO) accuracy. The only exception is the case of Drell-Yan production, where NNLO results are known for the total cross sections. So we generally have relatively small precision in the theoretical predictions, and theoretical uncertainties which are large when compared to LEP or HERA.

However, $p\bar{p}$ collider physics is primarily discovery physics, rather than precision physics (there are exceptions, such as the measurements of the $W$ mass and of the properties of $b$-hadrons. But these are not QCD-related measurements). As such, knowledge of QCD is essential both for the estimate of the expected signals, and for the evaluation of the backgrounds. Tests of QCD in $p\bar{p}$ collisions confirm our understanding of perturbation theory, or, when they fail, point to areas where our approximations need to be improved. (see, e.g., the theory advances prompted by the measurements of $\psi$ production at CDF).
Finally, a reliable theoretical control over the details of production dynamics allows one to extract important information on the structure of the proton (parton densities) in regions of $Q^2$ and $x$ otherwise unaccessible. Control of QCD at the current machines (the Tevatron at Fermilab) is therefore essential for the extrapolation of predictions to higher energies (say for applications at the future LHC, at CERN).

The key ingredients for the calculation of production rates and distributions in hadronic collisions are:

- the matrix elements for the hard, partonic process (e.g., $gg \rightarrow gg, gg \rightarrow b\bar{b}, q\bar{q} \rightarrow W, \ldots$),
- the hadronic parton densities, discussed in the previous lecture.

Then the production rate for a given final state $H$ is given by a factorization formula similar to the one used to describe DIS:

$$d\sigma(p\bar{p} \rightarrow H + X) = \int dx_1 dx_2 \sum_{i,j} f_i(x_1, Q) f_j(x_2, Q) d\hat{\sigma}(ij \rightarrow H),$$

(187)

where the parton densities $f_i$’s are evaluated at a scale $Q$ typical of the hard process under consideration. For example $Q \approx M_{DY}$ for production of a Drell-Yan pair, $Q \approx E_T$ for high transverse-energy ($E_T$) jets, $Q^2 \approx p_T^2 + m^2$ for high-$p_T$ heavy quarks, etc.

In this lecture we will briefly explore two of the QCD phenomena currently studied in hadronic collisions: Drell-Yan, and inclusive jet production. More details can be found in Refs. [8,4].

### 6.1. Drell-Yan processes

While the $Z$ boson has been recently studied with great precision by the LEP experiments, it was actually discovered, together with the $W$ boson, by the CERN experiments UA1 and UA2 in $p\bar{p}$ collisions. $W$ physics is now being studied in great detail at LEP2, but the best direct measurements of its mass by a single group still belong to $p\bar{p}$ experiments (CDF and D0 at the Tevatron). Even after the ultimate luminosity will have been accumulated at LEP2, with a great improvement in the determination of the parameters of the $W$ boson, the monopoly of $W$ studies will immediately return to hadron colliders, with the Tevatron data-taking resuming in the year 2000, and later on with the start of the LHC experiments.

Precision measurements of $W$ production in hadronic collisions are important for several reasons:

- this is the only process in hadronic collisions which is known to NNLO accuracy
- the rapidity distribution of the charged leptons from $W$ decays is sensitive to the ratio of the up and down quark densities, and can contribute to our understanding of the proton structure.
- deviations from the expected production rates of highly virtual $W$’s ($p\bar{p} \rightarrow W^* \rightarrow e\nu$) are a possible signal of the existence of new $W$ bosons, and therefore of new gauge interactions.

The partonic cross-section for the production of a $W$ boson from the annihilation of a $q\bar{q}$ pair can be easily calculated, giving the following result [8,4]:

$$\hat{\sigma}(q_i\bar{q}_j \rightarrow W) = \pi \frac{\sqrt{2}}{3} |V_{ij}|^2 G_F M_W^2 \delta(\hat{s} - M_W^2) = A_{ij} M_W^2 \delta(\hat{s} - M_W^2),$$

(188)

where $\hat{s}$ is partonic center of mass energy squared, and $V_{ij}$ is the element of the Cabibbo-Kobayashi-Maskawa matrix. The delta function comes from the $2 \rightarrow 1$ phase space, which forces the center-of-mass energy of the initial state to coincide with the $W$ mass. It is useful to introduce the two variables

$$\tau = \frac{\hat{s}}{S_{had}} \equiv x_1 x_2,$$

(189)
where $S_{\text{had}}$ is the hadronic center of mass energy squared. The variable $y$ is called rapidity. For slowly moving objects it reduces to the standard velocity, but, contrary to the velocity, it transforms additively even at high energies under Lorentz boosts along the direction of motion. Written in terms of $\tau$ and $y$, the integration measure over the initial-state parton momenta becomes:

$$dx_1 dx_2 = d\tau dy.$$  

Using this expression and Eq. (188) in Eq. (187), we obtain the following result for the LO total $W$ production cross section:

$$\sigma_{\text{DY}} = \sum_{i,j} \frac{\pi A_{ij}}{M_W^2} \int_{\tau}^{1} dx f_i(x) f_j \left( \frac{\tau}{x} \right) \equiv \sum_{i,j} \frac{\pi A_{ij}}{M_W^2} \tau \mathcal{L}_{ij}(\tau),$$

where the function $\mathcal{L}_{ij}(\tau)$ is usually called partonic luminosity. In the case of $u\bar{d}$ collisions, the overall factor in front of this expression has a value of approximately 6.5 nb. It is interesting to study the partonic luminosity as a function of the hadronic CoM energy. This can be done by taking a simple approximation for the parton densities. Following the indications of the figures presented in the previous lecture, we shall assume that $f_i(x) \sim 1/x^{1+\delta}$, with $\delta < 1$. Then

$$\mathcal{L}(\tau) = \int_{\tau}^{1} dx \frac{1}{x^{1+\delta}} \left( \frac{x}{\tau} \right)^{1+\delta} = \frac{1}{\tau^{1+\delta}} \int_{\tau}^{1} dx = \frac{1}{\tau^{1+\delta}} \log \left( \frac{1}{\tau} \right)$$

and

$$\sigma_W \sim \tau^{-\delta} \log \left( \frac{1}{\tau} \right) = \left( \frac{S_{\text{had}}}{M_W^2} \right)^{\delta} \log \left( \frac{S_{\text{had}}}{M_W^2} \right).$$

The DY cross-section grows therefore at least logarithmically with the hadronic CM energy. This is to be compared with the behaviour of the $Z$ production cross section in $e^+e^-$ collisions, which is steeply diminishing for values of $s$ well above the production threshold. The reason for the different behaviour in hadronic collisions is that while the energy of the hadronic initial state grows, it will always be possible to find partons inside the hadrons with the appropriate energy to produce the $W$ directly on-shell. The number of partons available for the production of a $W$ is furthermore increasing with the increase in hadronic energy, since the larger the hadron energy, the smaller will be the value of hadron momentum fraction $x$ necessary to produce the $W$. The increasing number of partons available at smaller and smaller values of $x$ causes then the growth of the total $W$ production cross section.

A comparison between the best available prediction for the production rates of $W$ and $Z$ bosons in hadronic collisions, and the experimental data, is shown in Fig. 4. The experimental uncertainties will soon be dominated by the limited knowledge of the machine luminosity, and will exceed the accuracy of the NNLO predictions. This suggests that in the future the total rate of produced $W$ bosons could be used as an accurate luminometer.

It is also interesting to note that an accurate measurement of the relative $W$ and $Z$ production rates (which is not affected by the knowledge of the total integrated luminosity, that will cancel in their ratio) provides a tool to measure the total $W$ width. This can be seen from the following equation:

$$\Gamma_W = \frac{N_{\text{obs}}^{Z \to e^+e^-}}{N_{\text{obs}}^{W \to e^+\nu}} \left( \frac{\sigma_{W\pm}}{\sigma_Z} \right) \left( \frac{\Gamma_W}{\Gamma_{Z,e^+e^-}} \right) \Gamma_Z.$$

As of today, this technique provides the best measurement of $\Gamma_W$: $\Gamma_W = 2.06 \pm 0.06$ GeV, which is a factor of 5 more accurate than the current best direct measurements from LEP2.
Fig. 4: Comparison of measured (a) $\sigma \cdot B(W \to e\nu)$ and (b) $\sigma \cdot B(Z^0 \to e^+e^-)$ to 2-loop theoretical predictions using MRSA parton distribution functions. The UA1 and UA2 measurements and D0 measurements are offset horizontally by ± 0.02 TeV for clarity. In the inset, the shaded area shows the 1σ region of the CDF measurement; the stars show the predictions using various parton distribution function sets (1) MRSA, (2) MRSD0', (3)MRSD-', (4) MRSH and (5) CTEQ2M. The theoretical points include a common uncertainty in the predictions from choice of renormalization scale ($M_W/2$ to $2M_W$).

6.2. $W$ rapidity asymmetry

The measurement of the charge asymmetry in the rapidity distribution of $W$ bosons produced in $p\bar{p}$ collisions can provide an important measurement of the ratio of the u-quark and d-quark momentum distributions. Using the formulas provided above, you can in fact easily check as an exercise that:

$$\frac{d\sigma_{W^+}}{dy} \propto f_u^p(x_1) f_d^p(x_2) + f_d^p(x_1) f_u^p(x_2),$$

(194)

$$\frac{d\sigma_{W^-}}{dy} \propto f_u^p(x_1) f_d^p(x_2) + f_d^p(x_1) f_u^p(x_2).$$

(195)

We can then construct the following charge asymmetry (assuming the dominance of the quark densities over the antiquark ones, which is valid in the kinematical region of interest for $W$ production at the Tevatron):

$$A(y) = \frac{\frac{d\sigma_{W^+}}{dy} - \frac{d\sigma_{W^-}}{dy}}{\frac{d\sigma_{W^+}}{dy} + \frac{d\sigma_{W^-}}{dy}} = \frac{f_u^p(x_1) f_d^p(x_2) - f_d^p(x_1) f_u^p(x_2)}{f_u^p(x_1) f_d^p(x_2) + f_d^p(x_1) f_u^p(x_2)}.$$ 

(196)
Setting \( f_d(x) = f_u(x) \, R(x) \) we then get:

\[
A(y) = \frac{R(x_2) - R(x_1)}{R(x_2) + R(x_1)},
\]

(197)

which measures the \( R(x) \) ratio since \( x_{1,2} \) are known in principle from the kinematics: \( x_{1,2} = \sqrt{\tau} \exp(\pm y) \). The current CDF data provide the most accurate measurement to date of this quantity (see Ref. [8]).

### 6.3. Jet production

Jet production is the hard process with the largest rate in hadronic collisions. For example, the cross section for producing at the Tevatron \( \sqrt{s_{\text{had}} = 1.8 \text{ TeV}} \) jets of transverse energy \( E_T \lesssim 50 \text{ GeV} \) is of the order of a \( \mu \text{b} \). This means 50 events/sec at the luminosities available at the Tevatron. The data collected at the Tevatron so far extend all the way up to the \( E_T \) values of the order of 450 GeV. These events are generated by collisions among partons which carry over 50\% of the available \( p \bar{p} \) energy, and allow to probe the shortest distances ever reached. The leading mechanisms for jet production are shown in Fig. 5.

![Fig. 5: Representative diagrams for the production of jet pairs in hadronic collisions.](image)

The 2-jet inclusive cross section can be obtained from the formula

\[
d\sigma = \sum_{ijkl} dx_1 \, dx_2 \, f_1^{(H_1)}(x_1, \mu) \, f_j^{(H_2)}(x_2, \mu) \, \frac{d\tilde{\sigma}_{ij \rightarrow k+l}}{d\Phi_2} \, d\Phi_2,
\]

(198)

that has to be expressed in terms of the rapidity and transverse momentum of the quarks (or jets), in order to make contact with physical reality. The two-particle phase space is given by

\[
d\Phi_2 = \frac{d^3k}{2k^0(2\pi)^3} \, 2\pi \, \delta((p_1 + p_2 - k)^2),
\]

(199)

and, in the CM of the colliding partons, we get

\[
d\Phi_2 = \frac{1}{2(2\pi)^2} \, d^2k_T \, dy \, 2\delta(s - 4k^0^2),
\]

(200)

where \( k_T \) is the transverse momentum of the final-state partons. Here \( y \) is the rapidity of the produced

---

\textsuperscript{4}In practice one cannot determine \( x_{1,2} \) with arbitrary precision on an event-by-event basis, since the longitudinal momentum of the neutrino cannot be easily measured. The actual measurement is therefore done by studying the charge asymmetry in the rapidity distribution of the charged lepton.
parton in the parton CM frame. It is given by

\[ y = \frac{y_1 - y_2}{2}, \tag{201} \]

where \( y_1 \) and \( y_2 \) are the rapidities of the produced partons in the laboratory frame (in fact, in any frame). One also introduces

\[ y_0 = \frac{y_1 + y_2}{2} = \frac{1}{2} \log \frac{x_1}{x_2}, \quad \tau = \frac{\hat{s}}{\hat{S}_{\text{had}}} = x_1 x_2. \tag{202} \]

We have

\[ dx_1 dx_2 = dy_0 d\tau. \tag{203} \]

We obtain

\[ d\sigma = \sum_{ijkl} dy_0 \frac{1}{\hat{S}_{\text{had}}} f_i^{(H_1)}(x_1, \mu) f_j^{(H_2)}(x_2, \mu) \frac{d\sigma_{ij \to k+l}}{d\Phi_2} \frac{1}{2(2\pi)^2} 2 dy d^2k_T, \tag{204} \]

which can also be written as

\[ \frac{d\sigma}{dy_1 dy_2 d^2k_T} = \frac{1}{\hat{S}_{\text{had}} 2(2\pi)^2} \sum_{ijkl} f_i^{(H_1)}(x_1, \mu) f_j^{(H_2)}(x_2, \mu) \frac{d\sigma_{ij \to k+l}}{d\Phi_2}. \tag{205} \]

The variables \( x_1, x_2 \) can be obtained from \( y_1, y_2 \) and \( k_T \) from the equations

\[ y_0 = \frac{y_1 + y_2}{2}, \tag{206} \]

\[ y = \frac{y_1 - y_2}{2}, \tag{207} \]

\[ x_T = \frac{2k_T}{\sqrt{\hat{S}_{\text{had}}}}, \tag{208} \]

\[ x_1 = x_T e^{y_0} \cosh y, \tag{209} \]

\[ x_2 = x_T e^{-y_0} \cosh y. \tag{210} \]

For the partonic variables, we need \( \hat{s} \) and the scattering angle in the parton CM frame \( \theta \), since

\[ t = -\frac{\hat{s}}{2} (1 - \cos \theta), \quad u = -\frac{\hat{s}}{2} (1 + \cos \theta). \tag{211} \]

Neglecting the parton masses, you can show that the rapidity can also be written as:

\[ y = -\log \tan \frac{\theta}{2} \equiv \eta, \tag{212} \]

with \( \eta \) being usually referred to as pseudorapidity.

The leading-order Born cross sections for parton parton scattering are reported in Table 1. It is interesting to note that a good approximation to the exact results can be easily obtained by using the soft-gluon techniques introduced in the third lecture. Based on the fact that even at 90° \( \min(|t|, |u|) \) does not exceed \( s/2 \), and that therefore everything else being equal a propagator in the \( t \) or \( u \) channel contributes to the square of an amplitude 4 times more than a propagator in the \( s \) channel, it is reasonable to assume that the amplitudes are dominated by the diagrams with a gluon exchanged in the \( t \) (or \( u \)) channel. It is easy to calculate the amplitudes in this limit using the soft-gluon approximation. For example, the amplitude for the exchange of a soft gluon among a \( qq' \) pair is given by:

\[ (\lambda_{ij}^a)(\lambda_{kl}^a) 2p_\mu \frac{1}{l} 2p'_\mu = \lambda_{ij}^a \lambda_{kl}^a \frac{4p \cdot p'}{l} = \frac{2s}{l} \lambda_{ij}^a \lambda_{kl}^a. \tag{213} \]
The $p_{t\mu}$ and $p'_{t\mu}$ factors represent the coupling of the exchanged gluon to the $q$ and $q'$ quark lines, respectively (see Eq. (105)). Squaring, and summing and averaging over spins and colours, gives

$$\sum_{\text{colours,spin}} |M_{qq'}|^2 = \frac{1}{N^2} \left( \frac{N^2 - 1}{4} \right) \frac{4s^2}{t^2} = \frac{8s^2}{9t^2}. \quad (214)$$

Since for this process the diagram with a $t$-channel gluon exchange is symmetric for $s \leftrightarrow u$ exchange, and since $u \rightarrow -s$ in the $t \rightarrow 0$ limit, the above result can be rewritten in an explicitly $(s,u)$ symmetric way as

$$\frac{4s^2}{9} \cdot \frac{s^2 + u^2}{t^2}, \quad (215)$$

which indeed exactly agrees with the result of the exact calculation, as given in Table 1. The corrections which appear from $s$ or $u$ gluon exchange when the quark flavours are the same or when we study a $q\bar{q}$ process are small, as can be seen by comparing the above result to the expressions in the Table.

As another example we consider the case of $qq \rightarrow qq$ scattering. The amplitude will be exactly the same as in the $qq' \rightarrow qq'$ case, up to the different colour factors. A simple calculation then gives:

$$\sum_{\text{colours,spin}} |M_{qq}|^2 = \frac{9}{4} \sum |M_{qq'}|^2 = \frac{s^2 + u^2}{t^2}. \quad (216)$$

The exact result is

$$\frac{u^2 + s^2}{t^2} - \frac{4}{9} \frac{u^2 + s^2}{us}, \quad (217)$$

which even at 90°, the point where the $t$-channel exchange approximation is worse, only differs from this latter by no more than 25%.

As a final example we consider the case of $gg \rightarrow gg$ scattering, which in our approximation gives:

$$\sum |M_{gg}|^2 = \frac{9}{2} \frac{s^2}{t^2}. \quad (218)$$

By $u \leftrightarrow t$ symmetry we should expect the simple improvement:

$$\sum |M_{gg}|^2 \sim \frac{9}{2} \left( \frac{s^2}{t^2} + \frac{s^2}{u^2} \right). \quad (219)$$

<table>
<thead>
<tr>
<th>Process</th>
<th>( \frac{d\sigma}{d^2t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( qq' \rightarrow qq' )</td>
<td>( \frac{1}{2} \left[ \frac{1}{9} \left( \frac{s^2 + u^2}{t^2} + \frac{s^2 + t^2}{u^2} \right) - \frac{8s^2}{27u^2} \right] )</td>
</tr>
<tr>
<td>( qq \rightarrow qq )</td>
<td>( \frac{1}{2} \frac{4}{9} \left( \frac{s^2 + u^2}{t^2} + \frac{t^2 + u^2}{s^2} \right) - \frac{8s^2}{27s^2} )</td>
</tr>
<tr>
<td>( qq \rightarrow q'q' )</td>
<td>( \frac{1}{2} \frac{4}{9} \left( \frac{s^2 + u^2}{t^2} + \frac{t^2 + u^2}{s^2} \right) - \frac{8s^2}{27s^2} )</td>
</tr>
<tr>
<td>( \bar{q}q \rightarrow q\bar{q} )</td>
<td>( \frac{1}{2} \frac{4}{9} \left( \frac{s^2 + u^2}{t^2} + \frac{t^2 + u^2}{s^2} \right) - \frac{8s^2}{27s^2} )</td>
</tr>
<tr>
<td>( gg \rightarrow gg )</td>
<td>( \frac{1}{2} \frac{4}{9} \left( \frac{s^2 + u^2}{t^2} + \frac{t^2 + u^2}{s^2} \right) - \frac{8s^2}{27s^2} )</td>
</tr>
</tbody>
</table>

Table 1: Cross sections for light parton scattering. The notation is $p_1, p_2 \rightarrow t \ell, \tilde{s} = (p_1 + p_2)^2, \tilde{t} = (p_1 - k)^2, \tilde{u} = (p_1 - l)^2$. The $p_{t\mu}$ and $p'_{t\mu}$ factors represent the coupling of the exchanged gluon to the $q$ and $q'$ quark lines, respectively (see Eq. (105)). Squaring, and summing and averaging over spins and colours, gives
This only differs by 20% from the exact result at 90°.

Notice that at small $t$ the following relation holds:

$$
\hat{\sigma}_{gg} : \hat{\sigma}_{qq} : \hat{\sigma}_{qg} = \left( \frac{9}{4} \right) : 1 : \left( \frac{4}{9} \right).
$$

(220)

The $9/4$ factors are simply the ratios of the colour factors for the coupling to gluons of a gluon ($C_A$) and of a quark ($T_F$), after including the respective colour-average factors ($1/(N^2 - 1)$ for the gluon, and $1/N$ for the quark). Using Eq. (220), we can then write:

$$
d\sigma_{\text{hadr}} = \int dx_1 \, dx_2 \sum_{i,j} f_i(x_1) \, f_j(x_2) \, d\hat{\sigma}_{ij} = \int dx_1 \, dx_2 \, \hat{F}(x_1) \, \hat{F}(x_2) \, d\hat{\sigma}_{gg}(gg \to \text{jets}),
$$

(221)

where the object:

$$
F(x) = f_g(x) + \frac{4}{9} \sum_f [q_f(x) + \bar{q}_f(x)]
$$

(222)

is usually called the effective structure function. This result indicates that the measurement of the inclusive jet cross section does not allow in principle to disentangle the independent contribution of the various partonic components of the proton, unless of course one is considering a kinematical region where the production is dominated by a single process. The relative contributions of the different channels, as predicted using the global fits of parton densities available in the literature, are shown in Fig. 6.

![Fractional Contribution to $d\sigma/dE_T$ at $\eta=0$](image)

Fig. 6: Relative contribution to the inclusive jet-$E_T$ rates from the different production channels.

Predictions for jet production at colliders are available today at the next-to-leading order in QCD. A comparison between these calculations and the available data is given in Figs. 7 and 8.

At the Tevatron, jets up to 450 GeV transverse momentum have been observed. That is $x \gtrsim 0.5$ and $Q^2 \simeq 160,000$ GeV$^2$. This is a domain of $x$ and $Q^2$ not accessible to HERA. The current agreement between theory and data is at the level of 30% over 8 orders of magnitude of cross-section, from $E_T \sim 20$ to $E_T \sim 450$ GeV. The small deviation observed by CDF at high $E_T$ is under active investigation both experimentally and theoretically. It is still premature to say whether it can be a signal of new phenomena, or whether it is the result of our incomplete knowledge of the gluon density at large $x$. Either way, future higher-statistics measurements at the Tevatron will provide some important input on these fundamental questions. The resulting knowledge will enable theorists to reliably predict production rates for all interesting processes that will take place at the LHC.
\[ \Delta \sigma \frac{d^2 \sigma}{d \eta} \frac{d \sigma}{d E_T} \text{ fb/GeV} \]

CTEQ3M \( \mu = E_T^{\text{max}}/2 \)

Unsmeared Data

Fig. 7: Inclusive \( E_T \) spectra for central jets at the Tevatron

Jet \( E_T \) (GeV) |
--- | --- | --- | --- | --- | --- | --- | --- |
10  | -1  | 10  | 10  | 10  | 10  | 10  |
20  | 10  | 10  | 10  | 10  | 10  | 10  |
30  | 10  | 10  | 10  | 10  | 10  | 10  |
40  | 10  | 10  | 10  | 10  | 10  | 10  |
50  | 10  | 10  | 10  | 10  | 10  | 10  |
60  | 10  | 10  | 10  | 10  | 10  | 10  |
70  | 10  | 10  | 10  | 10  | 10  | 10  |
80  | 10  | 10  | 10  | 10  | 10  | 10  |
90  | 10  | 10  | 10  | 10  | 10  | 10  |
100 | 10  | 10  | 10  | 10  | 10  | 10  |

Sys. Error (%)

Fig. 8: Comparison of inclusive jets cross sections with QCD calculations at the Tevatron.

ACKNOWLEDGEMENT

It is a pleasure to thank the organizers of this School, for the successful efforts made to bring top-quality students together, and to provide a great environment for physics discussions and for a pleasant time as well.
References

Standard Textbooks:


Pedagogical Reviews


Review Articles


Historical Reviews


Abstract
The successes and shortcomings of the Standard Model are reviewed, with emphasis on the reasons motivating the need to extend it. The basic elements of grand unification and supersymmetry are described, exploring their phenomenological implications for gauge coupling unification, proton decay, fermion masses, neutrino physics, collider signatures, dark matter, rare decays and anomalous magnetic moments. The forthcoming generation of experiments will certainly expose these ideas to several essential tests.

1. THE STANDARD MODEL

The experimental success of the $SU(3) \times SU(2) \times U(1)$ Standard Model (SM) of the strong and electroweak forces can be considered as the triumph of the gauge symmetry principle to describe particle interactions. As we now briefly summarize, there are many different facts which have to be taken into account when searching for a deeper underlying theory, and these are:

1.1. The good:
The major success of the SM is that it accounts for essentially all present accelerator results. In particular:

- The most accurately known quantity in particle physics is the magnetic moment of the electron, with its ‘anomalous’ part being $a_e \equiv (g_e - 2)/2$, i.e. the normalized difference in the gyromagnetic ratio $g_e$ with respect to the classical Dirac value $g_e = 2$ (where the magnetic moment $\mu$ is related to the particle spin $s$ through $\mu \equiv gs(e/2m)$). The experimental and theoretical values are respectively [1]

$$a_e = \begin{cases} (115965218.7 \pm 0.4) \times 10^{-11} \text{ Exp.} \\ (115965214.0 \pm 2.8) \times 10^{-11} \text{ Th.} \end{cases}$$

(1)

The theoretical expression results from the computation of the loop corrections within the SM up to order $O(\alpha^4)$, including diagrams such as those depicted in Fig. 1. The one loop photon correction is the well known Schwinger term. Its prediction was actually one of the first major successes of QED, proving that radiative corrections could be made meaningful through renormalization and were indeed measurable. Higher order contributions include terms such as that involving the photon self energy in diagram (1.c) and the photon–photon scattering appearing in diagram (1.d). These two are actually the major sources of uncertainties for the magnetic moments of heavier leptons ($\mu$ or $\tau$), but for the electron the theoretical error is actually dominated by the uncertainty in the direct measurement of $\alpha$ using the Quantum Hall Effect. Indeed, one may use the theoretical expression for $a_e$ (as a fourth order polynomial in $\alpha$) to infer a ‘theoretically based’ value for the electromagnetic coupling, which has a smaller error than the one obtained from direct measurements, and is $\alpha^{-1} = 137.03599993(52)$. This of course is the value at low energies, and running it to the scale of the $Z$ boson, where for instance LEP measurements are done, one gets in the modified minimal subtraction ($\overline{MS}$) scheme

$$\alpha^{-1}(M_Z) = \left( \frac{\alpha}{1 - \Delta \alpha} \right)^{-1} = 127.934 \pm 0.027,$$

(2)

where $\Delta \alpha$ encodes the effects of the radiative corrections.
Fig. 1: Some loop contributions to the anomalous magnetic moments of charged leptons.

- The weak interactions are based on the spontaneous breaking $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$ (with associated coupling constants $g$, $g'$ and $e$ respectively) induced by the vacuum expectation value of the neutral component of the Higgs doublet field $\langle H^0 \rangle = 246$ GeV. Assuming that the left handed fermions belong to SU(2) doublets while the right handed chiralities to SU(2) singlets ensures that the charge current couplings of the $W$ boson violate parity maximally, and this gives rise to the well established $V - A$ theory. At low energies, $W$-boson mediated processes have an effective coupling with strength given by the Fermi constant, $G_F = \frac{g^2}{2M_W^2}$, and the measured muon lifetime leads to the value $G_F = 1.16637(1) \times 10^{-5}$ GeV$^{-2}$. Regarding the neutral currents, the $\gamma$ and $Z$ bosons are obtained rotating by an angle $\theta_W$ the neutral SU(2) gauge boson $W^3$ and the $U(1)_Y$ boson $B$, i.e.:

$$\left( \begin{array}{c} A_\mu \\ Z_\mu \end{array} \right) = \left( \begin{array}{cc} c_W & s_W \\ -s_W & c_W \end{array} \right) \left( \begin{array}{c} B_\mu \\ W^3_\mu \end{array} \right).$$

(3)

The photon is the gauge boson associated to the unbroken $U(1)_{em}$ (and hence remains massless) provided that

$$\tan \theta_W = \frac{g'}{g},$$

(4)

and its coupling to fermions is vectorial, having strength $e \equiv g s_W$ and being proportional to the fermion charge $Q_f = T_3(f) + Y(f)$. Regarding the fermion couplings to the $Z$ boson, they can be written as $(g/c_W) \gamma_\mu (g_V^f - g_A^f \gamma_5)$, with the vectorial part being $g_V^f = T_3(f) - 2Q_f s_W^2$ and the axial vector piece being $g_A^f = T_3(f)$. The resonant production of $Z$ bosons at LEPI allowed to test these couplings and to accurately measure the $Z$ boson mass from the observed line-shape, resulting in $M_Z = 91.1872 \pm 0.0021$ GeV. One can also obtain directly the masses of the gauge bosons from the spontaneously broken electroweak Lagrangian and one obtains the relation $M_W = c_W M_Z$. Clearly this relation and Eq. (4) cannot both hold at the loop level, since the couplings in Eq. (4) run and are hence scale dependent. This leads to different definitions for the weak mixing angle, with the on-shell value being

$$\sin^2 \theta_W = 1 - \frac{M_W^2}{M_Z^2} = 0.22302 \pm 0.00040,$$

(5)

while the (scale dependent) $\overline{MS}$ one being

$$\sin^2 \hat{\theta}_W(\mu) = \frac{g'^2(\mu)}{g^2(\mu) + g'^2(\mu)},$$

(6)

with fits to electroweak observables leading to $\sin^2 \hat{\theta}_W(M_Z) = 0.23117 \pm 0.00016$. The $\overline{MS}$ definition of the mixing angle is the most appropriate one for the study of the running of gauge

\footnote{We denote $s_W \equiv \sin \theta_W$ and $c_W \equiv \cos \theta_W$.}
couplings, and in particular to confront unification predictions. Combining the large amount of
electroweak observables, including LEPI Z resonance cross sections, widths and asymmetries,
Tevatron and LEPII W-boson mass measurements, and also neutrino scattering processes, one can
test the effects of radiative corrections, which are sensitive to the virtual effects of the top quark
and Higgs boson. From these one obtains favored ranges for the top and Higgs masses, which
are $m_t = 174.1^{+9.7}_{-7.6}$ GeV and $m_H = 86^{+48}_{-32}$ GeV [2], in remarkable agreement with the top mass
measured at the Tevatron, $m_t = 174.3 \pm 5.1$ GeV, and suggesting the presence of a light Higgs boson\(^2\). Hence, one can say that also the electroweak sector has been tested at the loop level.

- The strong interactions are described by Quantum Chromo Dynamics, i.e. by the unbroken non-
abelian gauge theory SU(3). Due to the gluon’s self interactions it has the property of being
asymptotically free, while at low energies the coupling constant becomes large and the theory
should then account for the confinement of quarks into colorless hadrons. The running of the
strong coupling has been tested extracting $\alpha_s$ from experiments performed at different energies,
such as the measured $\tau$ lepton widths, deep inelastic scattering, Upsilon decays and $e^+ e^-$ colliders
at different center of mass energies and up to $\sim 200$ GeV. In particular, one has $\alpha_s(M_Z) =
0.119 \pm 0.003$.

- Besides the gauge sector, a crucial ingredient of the SM is the family structure. The first generation
of fermions consists of

$$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad e_R, \quad Q = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad u_R, \quad d_R$$

and this pattern is replicated two more times to lead to the three fermion families. This number of
three is nicely consistent with the number of massless neutrinos coupling to the Z-boson inferred
from the invisible Z width, $N_\nu = 2.994 \pm 0.012$, and also with the possibility of having CP
violation in the quark sector through a non-trivial phase in the Cabibbo Kobayashi Maskawa matrix
$V_{CKM}$. The CKM phase can account for the CP violating effects observed in the Kaon system ($\epsilon$
and $\epsilon'/\epsilon$) or B system (time dependent asymmetries in $B^0 \rightarrow J/\Psi K_S$).

The two quark states contained in the doublet $Q$ are directly coupled among them through a W
boson, and hence constitute the so-called flavor eigenstates. The mass eigenstate quarks are instead
in general a mixture of flavor eigenstates belonging to different families. Adopting for convenience
the up type ($u, c$ and $b$) flavor eigenstates to coincide with the corresponding mass eigenstates, the
down type mass eigenstates ($d' = (d', s', b')$) are just related to the flavor eigenstates through the
CKM matrix, i.e.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{7}$$

The unitarity of this matrix, which is ultimately due to the family structure of the model, ensures
that the couplings of the Z-boson to the down type mass eigenstates is also flavor diagonal, since
they turn out to be proportional to $(V^\dagger V)_{ij} = \delta_{ij}$. This is the basis of the Glashow Illiopoulos and
Maiani (GIM) mechanism [3], which forbids tree level flavor changing neutral currents (FCNC)
and suppresses the loop mediated ones, which are non-zero only due to the mass differences between
different quarks. This naturally accounts for the smallness of e.g. the $\Delta S = 1$ decay $K_L^0 \rightarrow \mu^+ \mu^-$
(with $BR \simeq 7 \times 10^{-9}$) or the $\Delta S = 2$ neutral kaon mass difference ($\Delta m_K/m_K \simeq 7 \times 10^{-15}$),
as illustrated in Fig. 2. The value of $\Delta m_K$ actually provided the first indication of the correct
mass for the charm quark [4], so that as in the previously mentioned example of the top mass, it
is important to keep in mind that in these two cases virtual processes were sensitive to the effects
of particles prior to their direct production at accelerators. One should then not be too skeptical
when looking for virtual effects as a way to discover new particles predicted in extensions of the

\(^2\)Some hints in favor of a Higgs mass of 115 GeV have even been suggested by LEPII data.
SM (e.g. supersymmetric or GUT ones).

Due to the large hierarchy $m_t \gg m_c$, in the flavor changing processes involving the $b$ quark, such as $b \to s\gamma$, the GIM suppression is not very effective, leading to sizeable decay rates. One has indeed

$$BR(B \to X_s\gamma) = \begin{cases} (3.72 \pm 0.33) \times 10^{-4} \text{ Th} \\ (2.96 \pm 0.35) \times 10^{-4} \text{ Exp} \end{cases}, \quad (8)$$

and the reduced suppression in this process makes it also an interesting probe for new physics.

1.2. The bad:

Together with the many good things of the SM, there are several bad aspects where it fails to provide an adequate solution. These are:

- There is by now quite solid evidence supporting the existence of non-zero neutrino masses, coming from the explanation of the atmospheric and solar neutrino problems in terms of neutrino oscillations. The SM with just left handed neutrinos and a Higgs doublet is unable to provide a non-zero neutrino mass. On the other hand, the masslessness of the neutrinos in the SM is not related to any deep symmetry principle (unlike the masslessness of the photon which is linked to gauge invariance), and hence it is quite common that extensions of the SM give rise to massive neutrinos. The simplest way to get massive neutrinos would be to introduce the right-handed neutrino states $\nu_R$, and obtain a Dirac mass term by means of a Yukawa coupling $L_Y = L \cdot R \cdot H + h.c.$ However, for the resulting masses to be sufficiently small (below the eV range) the Yukawa coupling would have to be quite small ($\lambda_\nu < 10^{-11}$), and this doesn’t seem very natural. All other attempts to give masses to the neutrinos require to extend the SM in more radical ways.

- The SM suffers from the so-called strong CP problem, which is the fact that a Lorentz invariant term in the SM Lagrangian of the form

$$L_{QCD} = \lambda_\nu \bar{\nu}_L \nu_R H + h.c., \quad (9)$$

is consistent with the gauge symmetries of the model and there is hence no reason to omit it (where $\bar{G}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma} G_{\rho\sigma} / 2$ is the dual of the gluon field strength $G_{\mu\nu}$). This term is however CP violating and induces a contribution to the neutron electric dipole moment $d_n \approx 5 \times 10^{-16} \theta_{QCD} e \text{ cm}$. The experimental upper bounds on $d_n$ require then that $\theta_{QCD} < 10^{-10}$, and the smallness of this parameter has no natural explanation within the SM. One possibility is to extend the Higgs sector and introduce an axial global U(1) symmetry, the so-called Peccei Quinn symmetry [5], which gets broken spontaneously leading to the appearance of a pseudo-Goldstone boson, the axion. The
An important theoretical difficulty of the SM is the hierarchy problem, related to the fact that loop corrections to the scalar (Higgs) masses are quadratically divergent, and hence in a regularized theory with an ultraviolet cutoff $\Lambda$ one gets contributions at one loop $\delta m_{H}^{2} \propto (1/16\pi^{2})\Lambda^{2}$, with the different contributions (see Fig. 3) being proportional to $\lambda$ (the Higgs quartic self-coupling), $g^{2}$ (the squared gauge coupling) or $-\lambda_{f}^{2}$ (the squared Yukawa coupling, with a minus sign due to the fermionic nature of the loop). If one thinks of the cutoff as the scale where a more fundamental (and less divergent) theory enters to play a role, it would be hard to understand how things conspire to cancel the large loop correction (if $\Lambda \sim M_{Pl}$ or $M_{\text{string}}$), leaving a Higgs mass at the TeV scale. Since this would require an accurate fine-tuning, the situation is usually referred to as the naturalness or fine-tuning problem. The simplest solution for this is to appeal to supersymmetry entering into the game at a relatively low scale ($\Lambda_{\text{SUSY}} \sim \text{TeV}$) and enforcing the cancellation between the dangerous quadratic divergences arising from bosonic and fermionic loops.

- The SM provides no explanation for the origin of the baryon asymmetry, i.e. the excess of matter over antimatter in the Universe. Even if there are baryogenesis scenarios exploiting the non-perturbative SM baryon number violation, in order for them to be successful new sources of CP violation and extended Higgs sectors are generally required.

- The SM has no good candidate for the dark matter which is inferred to contribute sizeably to the overall matter in the Universe. The favored cosmological model nowadays suggests indeed $\Omega_{\text{CDM}} \approx 0.3$ and $\Omega_{\Lambda} \approx 0.7$, and neither a Cold Dark Matter particle nor an explanation for the origin of a cosmological constant $\Lambda$ of the required size can be found within the SM. Cosmological theories of primordial inflation also require to search for their causes beyond the SM.

- An even more serious drawback is that the SM makes no attempt to include a consistent quantum theory of gravity.

1.3. The ugly:

The search for a good theory to describe particle interactions has also an esthetical component, and in physics beauty is generally related to simplicity and to the fact that important things should not happen by chance, but should instead result from solid underlying reasons. Some of the ugly things in the SM are:

- The model has many unrelated parameters. These include the three coupling constants $g_{3}$, $g$, and $g'$ and the Yukawa couplings, or equivalently three charged lepton masses, 6 quark masses, three CKM mixing angles and one CP violating phase. In the Higgs sector there is the Higgs quartic coupling and the Higgs VEV (or equivalently $M_{H}$ and $M_{W}$). There is also the QCD parameter $\theta_{\text{QCD}}$. This makes a total of 19 independent parameters. This number is further increased if we take into account the neutrino masses and leptonic mixings. Clearly a theory relating the gauge couplings (unification) or explaining the pattern of fermion mixings would be most welcome.

- The gauge group $SU(3) \times SU(2) \times U(1)$ was just put in by hand to explain observations, but there is no deep principle behind that choice.
• Assigning left-handed chiralities to SU(2) doublets and right-handed ones to singlets is again arbitrary. Left right symmetric models are believed to be more esthetic.
• The number of generations \( N_g = 3 \) is also unexplained.
• The quantization of electric charge, i.e. the fact that \( Q_d = Q_e / 3 \), is unexplained.
• The cancellation of gauge charge in the SM happens just by chance. Let us now further comment into the anomaly issue.

1.4. Anomalies

Anomalies occur when a classical symmetry of the Lagrangian is violated by quantum effects. A traditional example being the dilatation symmetry of a massless (i.e. scale free) theory, which gets broken at the quantum level due to the need to introduce the renormalization scale. This leads to a non-vanishing trace of the energy momentum tensor, which constitutes the so-called trace anomaly.

Another example are the chiral symmetries, i.e. a symmetry under transformations distinguishing between left and right fermion chiralities,

\[
\Psi \to e^{i \theta \gamma_5} \Psi \quad \left\{ \begin{array}{c}
\Psi_L = e^{-i \theta} \Psi_L \\
\Psi_R = e^{i \theta} \Psi_R
\end{array} \right.
\]

For a massless theory this is a good symmetry, and hence the associated Noether current \( J^5_{\mu} = \bar{\Psi} \gamma_{\mu} \gamma_5 \Psi \) is conserved. However, computing the one loop contribution arising from the triangle diagrams involving one axial vector and two vector couplings (and contracting this with the momentum incoming in the axial vector vertex, so as to get the Fourier version of the divergence of the current), one gets

\[
\partial^\mu J^5_\mu = \frac{g^2}{8\pi^2} F_{\mu \nu} F^{\mu \nu},
\]

with \( F_{\mu \nu} \) the field strength of the field coupling to the vector-like vertex. For the non-abelian case the resulting anomaly is proportional (as can be seen from Fig. 4) to the symmetric structure constants \( d_{abc} \equiv \text{Tr}[T_a \{ T_b, T_c \}] / 2 \), where \( T_i \) are the generators in the representation of the fermions running in the loop.

![Fig. 4: Diagrams contributing to the non-abelian anomaly.](image)

When the chiral symmetry in question is a global one, anomalies pose no problems, and may even be a blessing. This is the case for instance in massless QCD, where taking the limit \( m_u \simeq m_d \simeq m_s \simeq 0 \) the theory acquires an \( U(3)_L \times U(3)_R \) global symmetry, corresponding to separate unitary rotations of the left and right quark fields \( (u, d, s)_{L,R} \). This symmetry can be decomposed as \( SU(3)_V \times SU(3)_A \times U(1)_V \times U(1)_A \). Within QCD, the vectorial parts remain good symmetries, and are reflected in the baryon
multiplet degeneracies and in the baryon number conservation respectively, while $SU(3)_A \equiv SU(3)_{L-R}$ is spontaneously broken by the QCD chiral condensate, what is reflected in the appearance of Goldstone bosons which are just the meson octet. This chiral symmetry is anomalous and the anomaly is responsible for instance for the dominant $\pi^0$ decay ($\pi^0 \rightarrow \gamma\gamma$). The $U(1)_A$ symmetry is also anomalous, and this anomaly (combined with the non-trivial structure of the QCD vacuum) is responsible for giving a sufficiently large mass to the $\eta'$ meson. When considering the electroweak interactions, since both $SU(2)_L$ and $U(1)_Y$ are chiral gauge groups, i.e. have different couplings to left and right fermions, also the global vectorial symmetries of baryon and lepton number become anomalous, with

$$\partial^\mu J^B_\mu = \partial^\mu J^L_\mu = \frac{N_f}{32\pi^2} \left[ g^2 W_{\mu\nu} \tilde{W}^{\mu\nu} - g^2 Y_{\mu\nu} \tilde{Y}^{\mu\nu} \right],$$

where $N_f = 6$ is the number of flavors while $W_{\mu\nu}$ and $Y_{\mu\nu}$ are the corresponding field strengths. This anomaly implies that when topology changes in the gauge fields occur, baryon and lepton numbers will change by an amount $\Delta B = \Delta L = \Delta N_{CS}$, where the Chern Simons number $N_{CS}$ characterizes the amount of ‘winding’ of the gauge fields. These transitions are however very suppressed at low energies/temperatures, since they are mediated by instantons, but become however very efficient at high temperatures, where they take place through ‘sphaleron’ excitations, and this has profound implications for baryogenesis theories.

Although global anomalies are welcome, when an anomaly affects a gauge symmetry this can be a disaster, since the gauge symmetry principle is at the basis of the formulation of the theory and is essential for its renormalizability. Hence, if the gauge symmetry ceases to be valid at the loop level one will certainly be in trouble. Due to the chiral anomaly, this will in general be the case for a chiral gauge theory, such as the one present in the electroweak model, so that the renormalizability of the SM is threatened. However, it turns out that when one adds the contributions to the anomaly coming from the leptons and from the $N_c = 3$ colors of quarks running in the triangle diagram, one has luckily a fortuitous cancellation, since

$$\sum \left( \text{LEPTON ANOMALY} + N_c \times \text{QUARK ANOMALY} \right) = 0.$$  

(13)

For instance, the $SU(2)_L^2 \times U(1)_Y$ contribution is proportional to (using $Y = Q - T_3$, $\{T_a, T_b\} = \delta_{ab}/2$ and $\text{Tr} T_3 = 0$)

$$\text{Tr} \left[ \{T_a, T_b\} Y \right] = \frac{\delta_{ab}}{2} \sum_i Q_i \propto \left[ Q_c + 3(Q_d + Q_u) \right] = 0.$$  

(14)

It is clear then that the SM with quarks or leptons alone would be anomalous, and to get a consistent electroweak theory we need to have both of them simultaneously. A theory in which anomalies are absent independently of the choice of matter representation (such as $SO(10)$ GUTs), would not have this item in the list of ugly things.

### 1.5. The road to unification

Many of the major advances in physics have resulted from the unified description of aspects which were before believed to be unrelated. This is also an important guide in the search of the underlying theory behind the SM, trying to remedy its pitfalls without loosing its successes. A schematic history of these unifications is illustrated in Table 1, and we will discuss hereafter some aspects of unification and supersymmetry (but not deal however with other also fundamental ideas such as extra dimensions or strings).
Table 1: Brief history of unifications.

2. GRAND UNIFIED THEORIES

The idea in Grand Unified Theories (GUTs) is to embed the SM group $SU(3) \times SU(2) \times U(1)$ in a larger group, with the aim of relating the different gauge couplings and also hoping that the mass spectrum will become simpler, since the quarks and leptons become different aspects of the same GUT ‘particle’. The most famous examples of unified groups are

- The Pati–Salam model [6] was the first GUT theory proposed (1972). It was based on the $SU(4) \times SU(2)_L \times SU(2)_R$ symmetry, with the leptons being some kind of fourth quark flavor and each family consisting of the representations

$$\begin{pmatrix} u_1 & u_2 & u_3 & \nu_\ell \end{pmatrix}, \begin{pmatrix} d_1 & d_2 & d_3 & \ell \end{pmatrix}_{L,R}.$$  

(15)

Putting quarks and leptons in the same $SU(4)$ multiplets lead to proton decay processes, and the left-right symmetry required the introduction of right handed neutrinos in each generation. Although each group factor has an associated gauge coupling, relations among them can be imposed invoking discrete symmetries.

- The smallest (i.e. rank = 4) simple gauge group containing the SM is $SU(5)$, and it was studied by Georgi and Glashow in 1974 [7]. Each generation is contained in two $SU(5)$ irreducible representations, which are \(^3\)

$$10 : \begin{pmatrix} 0 & u_2^c & -u_1^c & u_1 & d_1 \\ -u_3 & 0 & u_1^c & u_2 & d_2 \\ u_2 & -u_1^c & 0 & u_3 & d_3 \\ -d_1 & -u_2 & -u_3 & 0 & e^c \\ -u_1 & -u_2 & -u_3 & 0 & e^c \end{pmatrix}, \ \ \ \ \ \ \ 5 : \begin{pmatrix} d_1^c \\ d_2^c \\ d_3^c \\ e \\ -\nu \end{pmatrix}.$$  

(16)

- The orthogonal group $SO(10)$ is the next choice. It has rank = 5 and can contain the previously mentioned GUT groups. The fermions of each generation are contained in just one irreducible representation, the 16, which can be decomposed under $SU(5)$ as $16 = 10 + 5 + 1$, with the singlet state being the right handed neutrino.

Another attractive aspect of SO(10) is that orthogonal groups are automatically anomaly free, since their symmetric structure constants $d_{abc}$ vanish. This is not the case in SU(5), where one still needs the fortuitous cancellation between the non-vanishing anomalies coming from the 5 and the 10, which just happen to be opposite.

\(^3\)In GUTs it is always convenient to use the left handed conjugate fields $(\Psi^c)_L$ rather than the right handed fields $\Psi_R$, and they clearly describe the same degrees of freedom, since $(\Psi_R)^c \equiv C \Psi_R^T = (\Psi^c)_L$. 

92
• Other larger groups which have been intensively studied are: the exceptional group $E_6$ (with fermions in the $27$, which under $SO(10)$ decomposes as $16 + 10 + 1$, so that many exotic states appear in each generation); the string inspired groups $E_8 \times E_8$ (with $E_6$ being contained in one of the $E_8$ factors) and $SO(32)$. Large symmetry breaking chains are required to go from these large groups down to the SM, and many new heavy particles are left around when doing so (heavy Higgses, new fermions, new gauge bosons such as $W_R$ or $Z'$ ones), and these are the object of many dedicated searches at accelerators.

2.1. Gauge coupling unification

In the SM the three gauge couplings are different and unrelated, i.e. $g_3 \neq g \neq g'$. Moreover, the actual value of $g'$ is related to the arbitrary normalization of the hypercharge generator. Indeed, consider the covariant derivative in the electroweak model, which is

$$D_\mu = \partial_\mu - igT^a_\mu W^a - ig'Y_1 B_\mu.$$  

(17)

The gauge fields are clearly normalized through their kinetic terms, while the $SU(2)$ generators are normalized through the non-linear condition

$$\text{Tr}(T^a T^b) = \frac{\delta_{ab}}{2},$$

(18)

and this fixes the normalization of the coupling $g$. On the other hand, the $U(1)_Y$ coupling $g'$ is arbitrarily fixed by adopting the hypercharge normalization from the relation $Y = Q - T^3$. In an unified gauge group where all generators are similarly normalized, the hypercharge generator will turn out to have a different normalization than in the SM, and this modifies the expected relation between the GUT coupling $g_{GUT}$ and $g'$. Take for instance the case of $SU(5)$, where in the fundamental representation the 24 generators, normalized through the relation (18), can be conveniently written in block form using the $SU(3)$ Gellman matrices $\lambda^i$ and $SU(2)$ Pauli matrices $\sigma^i$, as

$$T^i = 1, \ldots, 8 = \frac{1}{2} \left( \begin{array}{cc} \lambda^i & 0 \\ 0 & 0 \end{array} \right), \quad T^{9,10,11} = \frac{1}{2} \left( \begin{array}{cc} 0 & 0 \\ 0 & \sigma^{1,2,3} \end{array} \right)$$

$$T^{12} = \sqrt{\frac{3}{5}} Y, \quad T^{13, \ldots, 24} : \text{off diagonal}$$

(19)

We see that $T^{12}$ is proportional to the hypercharge generator $Y = \text{diag}(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, but is now properly normalized. The covariant derivative of the unified group is obtained as $D_\mu = \partial_\mu - ig_{GUT} V^a_\mu T^a = \partial_\mu - ig_{GUT} \left( + W^a_\mu T^{11} + B_\mu T^{12} + \ldots \right)$. From this we see that the correct identification should be

$$g = g_{GUT}, \quad g' = \sqrt{\frac{3}{5}} g_{GUT}.$$  

(20)

This clearly implies that

$$\sin^2 \theta_W = \frac{g'^2}{g^2 + g'^2} = \frac{3}{8}.$$  

(21)

This prediction, which corresponds to $s_W^2 = 0.375$, is clearly far from the measured value, but this can be remedied once it is realized that it should hold at the GUT scale, which for many reasons turns out to be quite large ($M_{GUT} > 10^{14}$ GeV), and hence the prediction is sizeably affected by the running of the gauge couplings.
Notice that the photon should be contained among the GUT gauge bosons, and indeed the charge generator is just given by \[ Q = T^{11} + \sqrt{5/3}T^{12}. \] Since the generators are traceless, this implies that \[ \text{Tr} Q = 0. \] Considering for instance the fundamental \( \bar{5} \) representation, this implies that \( Q_e = 3Q_d \), explaining the charge quantization relation just from the fact that quarks and leptons are inside the same GUT multiplets.

2.2. Running of gauge coupling

To obtain the running of the gauge couplings in the SM it is necessary to consider the loop corrections to the vertex and wave functions of the gauge bosons (depicted in Fig. 5), or alternatively one may consider the gauge boson coupling to fermions and the fermion wave function renormalization. The result of these two approaches is of course the same, as expressed in a Slavnov Taylor identity reflecting the constraints imposed by the gauge symmetry.

\[\begin{split}
\text{VERTEX} &+ \\
\text{WAVE} &+ \
\end{split}\]

Fig. 5: Diagrams determining the vertex and gauge boson wave function renormalization, required to compute the beta functions. The scalar lines correspond to Fadeev-Popov ghosts, and in the case of SU(2) and \( U(1)_Y \) include the Higgs bosons.

The evolution of the coupling constant is obtained from the solution of the renormalization group equations (RGE) \( \frac{dg}{dt} = \beta \), with \( t \equiv \ln(\mu/\mu_0) \) specifying the momentum scale. For the SU(N) group one obtains the beta function from direct computation of the diagrams in Fig. 5 (e.g. they are obtained at one loop from the residues of the \( 1/\varepsilon \) poles in dimensional regularization), and they are

\[
\beta = \frac{g^3}{16\pi^2} \left[ -\frac{11}{3} t_2(V) + \frac{2}{3} t_2(F) + \frac{1}{3} t_2(S) \right],
\] (22)

with \( t_2 \) defined through

\[
\text{Tr} \left[ T^a T^b \right] = t_2 \delta^{ab},
\] (23)

and where the generators are in the appropriate representation of vectors \( (V) \), fermions \( (F) \) or scalars \( (S) \). For the fields in the adjoint one has that the Casimir is \( t_2(V) = N \), while for the fundamental representation one has \( t_2(F, S) = 1/2 \) for either Weyl fermions or complex scalars (and twice as much for a Dirac fermion). Hence, for QCD one has

\[
\beta_{QCD} = \frac{g^3}{16\pi^2} \left[ -11 + \frac{2}{3} \times \frac{1}{2} \times N_f \times 2 \right],
\] (24)

including the \( N_f = 6 \) flavors of Dirac quarks coupled to the gluons. It proves convenient to introduce
the factors $b_i$ such that
\[ \frac{dg_i}{dt} = \frac{b_i}{16\pi^2} g_i^3, \quad (25) \]
and hence the solution of the RGE for the running coupling constant at one loop can be expressed as
\[ g_i^2(t) = \frac{g_i^2(0)}{1 - (g_i^2(0)/8\pi^2)b_i t}. \quad (26) \]

For the SU(3), SU(2) and U(1) couplings of the SM one then gets (using also that for $U(1)_Y$ one has $t_2(F, S) = Y^2$ while $t_2(V) = 0$)
\[ b_3 = -7, \quad b_2 = -\frac{19}{6}, \quad b_1 = \frac{41}{10}. \quad (27) \]

We can then start from the measured values of the coupling constants at the scale $M_Z$, which are obtained from $\alpha_s$, $\alpha$ and $\sin^2\theta_W$ through $g_3 = \sqrt{4\pi\alpha_s}$, $g_2 \equiv g = \sqrt{4\pi\alpha/\sin\theta_W}$ and $g_1 = \sqrt{5/3}g \tan\theta_W$, and run them to high energies using the RGE. The result is plotted in Fig. 6. We see that the negative value of $b_3$ makes the strong coupling to become weaker at high energies, i.e. QCD becomes asymptotically free. Also $g_2$ becomes smaller in the SM with increasing energies, but with a smaller slope, while $g_1$ increases with energy\(^4\). We see that although the three couplings are quite different at low energies, they have the tendency to unify at a scale $M_X \approx 10^{14} – 10^{15}$ GeV. This is certainly encouraging and may be pointing to the existence of an underlying GUT symmetry at large scales. However, within the SM the convergence of the gauge couplings at one scale is not perfect (and this is not solved by including two loop contributions nor threshold effects), so that something more will be required to get an accurate unification of the couplings at a unique scale.

\[ 4 \text{In any case the Landau pole, i.e. the energy scale at which } g_1 \text{ would blow up according to Eq. (26), is well beyond the Planck scale.} \]
2.3. GUTs and fermion masses

Another important issue that GUTs can address is to explain some of the observed patterns of fermion masses. Let us consider for instance the case of SU(5) in some detail. The breaking of SU(5) down to the SM is performed by the VEV of an adjoint $24$ Higgs representation, so as to preserve the rank of the group. One has then $SU(5) \times SU(3) \times SU(2) \times U(1)$. Since the SM fermions can acquire mass only through the bilinears

$$
\begin{align*}
5 \times 5 &= 10 + \overline{10} \\
10 \times 10 &= 5 + 45 + 50 \\
5 \times 10 &= 5 + \overline{45}
\end{align*}
$$

it is clear that one cannot make a singlet by contracting these fermion bilinears with the $24$. Hence, SM fermions do not acquire masses in the SU(5) breaking stage (what is fortunate of course, since they would be otherwise too heavy). To give masses to both $5$ and $10$ representations with just one Higgs multiplet, one has then to use either the $5$ or the $45$ representations. The minimal SU(5) model uses just one $5$ of Higgses $\Phi$ to give fermions a mass in the process of electroweak symmetry breaking $SM \times SU(3) \times U(1)_{em}$. This Higgs representation consists of a color triplet state $T$ and the usual Higgs doublet $H$, i.e. $5 = (T, H)^T$. The Yukawa couplings can be written as

$$
\mathcal{L}_{Y} = f_{\alpha\beta} \chi_{\alpha ij} C \chi_{\beta kl} \Phi_{m} e^{ijklm} + f_{\alpha\beta} \chi_{\alpha ij} C \Psi_{\beta ij} \Phi_{m} + h.c.,
$$

where $\chi_{ij}$ are the fermions in the $10$ and $\Psi^{i}$ those in $\overline{5}$, while $\alpha, \beta$ are family indices. When $5$ acquires a VEV breaking the electroweak symmetry, i.e. $\langle \Phi \rangle = (0, 0, 0, 0, v)$, one gets masses for the fermions, and those of the down quarks and leptons satisfy

$$
M_{\beta\alpha}^{(d)} = M_{\alpha\beta}^{(f)} = v f_{\alpha\beta}^{(2)}.
$$

This relation would imply in particular the equality $m_{b} = m_{\tau}$, which again has to be checked at the GUT scale rather than at low energies. To take into account the running of fermion masses one has to evaluate the diagram in Fig. 7, which gets contributions from color and hypercharge gauge bosons in the loop (since $SU(2)_{L}$ bosons cannot provide the required chirality flip). From these one gets

$$
\frac{m_{b}(\mu)}{m_{\tau}(\mu)} = \left( \frac{g_{3}(\mu)}{g_{GUT}(M_{X})} \right)^{8/(11-2N_{f}/3)} \left( \frac{g_{1}(\mu)}{g_{GUT}(M_{X})} \right)^{3/N_{f}}.
$$

This gives at low energies ($\mu \simeq \text{GeV}$) the prediction $m_{b} \simeq 3m_{\tau}$, which is in rough agreement with the measured values (we have here neglected the effects of Yukawa couplings). This seems to be then another success of the GUT idea, but when looking however to the first two generations, the same relations would imply $m_{s} \simeq 3m_{\mu}$ and $m_{\mu}/m_{e} \simeq m_{s}/m_{d}$, which are however badly violated, since experimentally $m_{\mu} \simeq m_{s}$ and $m_{\mu}/m_{e} \simeq 10m_{s}/m_{d}$. One proposed solution to fix this is to include also

![Fig. 7: Diagram responsible for the running of the masses.](image-url)
a 45 Higgs representation to generate the fermion masses \([8]\). This one has the property of introducing a Clebsch Gordan factor of \(-3\) for the lepton mass term with respect to that of the down quarks, so that the prediction arising from the 45 induced mass term alone would be \(m_\ell = 3m_d\) at the GUT scale. Running this then down to low energies would lead to \(m_\ell \simeq m_d\), as seems to be the case for the second generation. Hence it would seem desirable to exploit the coupling to the 5 to generate the third generation masses, while that to the 45 to generate the second generation ones. Furthermore, with appropriate Yukawas (with ‘texture’ zeroes imposed by discrete symmetries) one may fix the masses of the three generations. For instance, taking

\[
\lambda_5 \langle 5 \rangle \sim \begin{pmatrix} 0 & A & 0 \\ A & 0 & 0 \\ 0 & 0 & B \end{pmatrix}, \quad \lambda_{45} \langle 45 \rangle \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & 0 \end{pmatrix}
\]

one gets, after taking into account the above mentioned Clebsch, that

\[
M_d \propto \begin{pmatrix} 0 & A & 0 \\ A & C & 0 \\ 0 & 0 & B \end{pmatrix}, \quad M_\ell \propto \begin{pmatrix} 0 & A & 0 \\ A & -3C & 0 \\ 0 & 0 & B \end{pmatrix}.
\]

This gives \(m_b = m_\tau \simeq B\), which is good at the GUT scale. For \(B \gg C \gg A\) one has also \(m_s \simeq C \simeq m_\mu/3\) and \(m_d \simeq A^2/C \simeq 3m_e\). When running down to low energies one obtains the satisfactory results \(m_b \simeq 3m_\tau, m_s \simeq m_\mu\) and \(m_\mu/m_e \simeq 9m_s/m_d\). Of course the ‘textured’ Yukawas in Eq. (32) were put in by hand, but the success of the predictions may be giving a hint on the kind of symmetries required to generate them.

### 2.4. GUTs and proton decay

In the SM there is no way to write renormalizable terms in the Lagrangian consistent with the gauge symmetries but violating baryon number \(B\). Hence, one has the fact that \(B\) is an accidental symmetry (it was not imposed by any deep fundamental reason) of the SM Lagrangian, and is only violated by the non-perturbative effects related to the anomaly mentioned before. As a consequence, the proton turns out to be stable in the SM, since it is the lightest particle with non-zero \(B\). When GUTs are considered, we have seen that quarks and leptons reside together in large GUT multiplets. This means that there will be gauge bosons connecting them, and hence violating \(B\) and \(L\). For instance, of the 24 gauge bosons in the adjoint of \(SU(5)\) there are, besides the 12 ones belonging to the SM, additional color triplet and weak doublets vector fields \((X_\mu^\alpha, Y_\mu^\alpha)\) (with \(\alpha = 1, 2, 3\) the color index), which together with their antiparticles make the twelve remaining gauge bosons. They have electric charges \(Q(X) = 4/3\) and \(Q(Y) = 1/3\), and couple to the fermions through

\[
\mathcal{L} \supset \frac{g_5}{\sqrt{2}} \left\{ X_\mu^\alpha \left[ \bar{d}_\alpha \gamma^\mu e + \bar{d}_\alpha \gamma^\mu e + \epsilon_{\alpha\beta\gamma} \bar{u}_\beta \gamma^\gamma \gamma^\mu u_\gamma \right] + Y_\mu^\alpha \left[ -\bar{d}_\alpha \gamma^\mu \nu - \bar{u}_\alpha \gamma^\mu e + \epsilon_{\alpha\beta\gamma} \bar{u}_\beta \gamma^\gamma \gamma^\mu d_\gamma \right] \right\}.
\]

Hence, we see that \(X_\mu\) couples as a ‘leptoquark’ \((X \to e^+ \bar{d})\) and as a ‘diquark’ \((X \to uu)\), clearly violating \(B\) and \(L\), although always preserving \(B - L = 2/3\). The same happens with the couplings of \(Y_\mu\) \((Y \to \bar{\nu}d, ud)\). Combining these couplings one can construct the proton decay diagrams shown in Fig. 8, which lead to \(p \to e^+ \pi^0, \bar{\nu} \pi^+\). There are similarly diagrams leading to \(n \to \bar{\nu} \pi^0\), although no \(n \to e^- \pi^+\) is allowed by \(B - L\) conservation. When contracting the heavy gauge boson propagator one gets an effective dimension six operator \(ggq_\ell M_X^2\) which is the one responsible for the \(p\)-decay. To compute the decay rate is somewhat delicate, since QCD corrections with gluons exchanged between the external quark lines are sizeable (overall factor \(\sim 3\)), and there are also significant uncertainties coming from the hadronic matrix elements, which have to be estimated from hadronic bag models or QCD sum
rules. As a result, one gets typical estimates for the proton lifetime

$$\tau(p \rightarrow e^+ \pi^0) \simeq (0.2 - 8) \times 10^{31} \left( \frac{M_X}{10^{15} \text{ GeV}} \right)^4 \text{ yr}. \quad (35)$$

Although this lifetime is huge, it is measurable by looking to large quantities of protons (1 kton of water has some $10^{33}$ protons). Actually the search for proton decay to test GUT predictions was the original purpose of the large underground detectors IMB and Kamiokande. However, these searches proved to be in vain, and the present bound set by Superkamiokande is $\tau(p \rightarrow e^+ \pi^0) > 1.6 \times 10^{33}$ yr, clearly excluding the simplest version of SU(5) GUT.

2.5. The doublet-triplet splitting problem

We have seen that in SU(5) the SM Higgs doublet belongs to a 5 in which also a triplet is present. This colored triplet also mediates proton decay through the diagram in Fig. 9. Although the diagram is suppressed by Yukawa couplings as compared to the gauge boson mediated ones, it would anyhow lead to extremely rapid proton decay unless the triplet states are sufficiently heavy, typically $m_T > 10^{12}$ GeV. Making the triplet state so heavy, while at the same time keeping the SM doublet Higgs sufficiently light ($m_H < \text{TeV}$) constitutes the so-called doublet-triplet splitting problem, which is one of the challenges for the GUT theoreticians. Writing the most general scalar potential with the Higgses present in minimal SU(5), i.e. with 5 and 24, gives rise to mass terms for both $H$ and $T$ of the order of the GUT scale. It is possible however to fine-tune the couplings in the potential so that $m_H = 0$ at tree level, but this relation is generally not stable under radiative corrections. One possible solution, the “missing partner” mechanism, is to couple the 5 to a multiplet containing triplet states but no doublets (such as the 50 in SU(5)), and hence through this coupling, which also involves the 75, only the triplet acquires a (large) mass, with the doublet remaining light. Another possibility is to start with a scalar potential with a large global symmetry (larger than the GUT gauge one), and such that when the symmetry is broken the doublet Higgs boson remains as a Goldstone boson of the spontaneously broken global symmetry [9], being then naturally light at tree level without the need of fine-tuning the couplings. Yet another possibility (and there are many more) which has received attention recently involves the GUT breaking by orbifold compactification of extra dimensions [10].
2.6. GUTs and $m_\nu$

Lacking any observation of proton decay, nucleon decay experimenters turned then to study just the background in their detectors, which consists of the atmospheric neutrinos produced by the cosmic rays hitting the top of the atmosphere. These lead to the greatest success of those experiments through the observed deficit in the ratio of $\nu_\mu/\nu_\tau$ fluxes. In particular, the zenith angle dependence of this ratio measured by the SuperKamiokande experiment lead to a very clean signal supporting neutrino oscillations (such that muon neutrinos oscillate into tau neutrinos in their way from the top of the atmosphere up to the detectors, and at multi-GeV energies can do that efficiently when coming from below but don’t have enough time to oscillate when coming from above). The implied neutrino mass difference $\Delta m^2 \simeq 3 \times 10^{-3} \text{eV}^2$ constitutes indeed the main evidence in favor of physics beyond the SM that we have at present.

The favored mechanism to generate naturally small neutrino masses is the so-called see-saw mechanism [11]. This requires the existence of right handed neutrinos, which being SM singlets can naturally acquire a large mass in some stage of GUT breaking. Through the combined action of a Dirac mass term and the singlet Majorana mass term, with

$$ -\mathcal{L}_m = \frac{1}{2} \begin{pmatrix} \nu_L^c \nu_R \end{pmatrix} \begin{pmatrix} 0 & m_D \cr m_D^* & M_R \end{pmatrix} \begin{pmatrix} \nu_L^c \nu_R \end{pmatrix} + h.c., \tag{36} $$

one gets the light mass eigenstates with masses (here we are ignoring the family structure)

$$ m_\nu \simeq \frac{m_D^2}{M_R} \simeq 10^{-3} \text{eV} \left( \frac{m_D}{\text{GeV}} \right)^2 \left( \frac{10^{12} \text{GeV}}{M_R} \right), \tag{37} $$

so that the lightness of the neutrinos is just due to the heaviness of the right handed states.

The necessary appearance of right handed neutrinos in GUTS such as SO(10) is then most welcome, and the see-saw mechanism naturally fits within those models (and in general in left-right GUT models also).

One difficulty however with SO(10) is that to give a mass to $\nu_R$ through a Yukawa coupling involving the fermion bilinear $16 \times 16 = 10 + 126 + 120$ requires the introduction of a large Higgs multiplet, the 126, since by decomposing the SO(10) multiplets above into their SU(5) content, it can be checked that only the 126 contains the singlet state corresponding to the bilinear right handed neutrino combination. Such large representations are however not found for instance when obtaining these GUT theories from the field theory limit of a superstring theory. One possibility to give mass to the right-handed neutrino in SO(10) without introducing a 126 is to do it radiatively with two vacuum insertions of a scalar 16 effectively acting as a 126, or more generally just using non-renormalisable couplings.

2.7. GUTs and baryogenesis

Very soon after the observation of CP violation in the Kaon system, Sakharov realized that it was possible to generate dynamically the baryon asymmetry through microphysical processes taking place in the early Universe. Besides the CP (and C) violation, which account for the asymmetry between particles and antiparticles, he found that it was necessary to have baryon number violating interactions (so as to have $B \neq 0$ at the end starting from an initial state with $B = 0$) and to be out of equilibrium, so that these same interactions do not erase the generated baryon asymmetry. At the time there were no theories predicting $B$-violating interactions, so that when GUT models were proposed to unify the gauge interactions, the prediction that they should lead to $B$-violating couplings was very welcome by cosmologists. Since this $B$-violation is linked to very heavy particles (the triplet Higgses $T$ or the $X,Y$ gauge bosons for the SU(5) case), the departure from equilibrium could take place just after these particles become non-relativistic in the very hot Universe. The annihilation and decay rates at this stage may not be fast enough
to keep the densities at their equilibrium values in the rapidly cooling Universe. The traditional scenario for baryogenesis was for instance to have the heavy SU(5) Higgs triplets decaying out of equilibrium and through $B$ and $CP$ violating decay channels. $CP$ violation implied e.g. $\Gamma(T \rightarrow uu) \neq \Gamma(T^* \rightarrow \bar{u}\bar{u})$, and this can occur once loop effects are considered.

However, when in the eighties it was realized that sphaleron mediated $B$ violation in the SM was unsuppressed at temperatures $> 100$ GeV [12], it was clear that this could have the effect of erasing any asymmetry produced previously at the GUT stage. Electroweak anomalous processes change both $B$ and $L$, but leaving intact $B - L$. Hence, the way out was to have a GUT theory generating initially a non-zero $B - L$, but this was not the case in the simplest SU(5) model, which preserves $B - L$. The other possibility would be to have baryogenesis at (or below) the electroweak scale, when sphalerons are no more active.

An interesting proposal is the so-called leptogenesis [13], in which the heavy right-handed neutrinos of the see-saw model decay at early times violating lepton number (with $CP$ violation resulting in $\Gamma(N \rightarrow LH) \neq \Gamma(N \rightarrow LH^*)$). The lepton asymmetry so produced is later partially transformed into a baryon asymmetry by the sphaleron processes. This is at present the most natural way of explaining the observed excess of matter over antimatter, since it just requires the see-saw mechanism and to be not particularly unlucky with the choice of model parameters.

### 3. SUPERSYMMETRY

#### 3.1. The naturalness problem

In general one says (following t’Hooft) that a small parameter in a theory is natural when setting it to zero increases the symmetry of the problem, so that this very same symmetry is the responsible for the smallness of the parameter. For instance, when applied to particle masses one has that the masslessness of a vector field can be related to the gauge invariance of the theory (as in the case of the photon), while the vanishing of a Dirac fermion mass is associated to a chiral symmetry (and the vanishing of a neutrino Majorana mass may be associated to a lepton number symmetry). However, for a scalar field in general no symmetry is gained when setting the mass to zero, except in the very particular case in which the boson is the Goldstone boson of a global symmetry. This implies that even if we set by hand the tree level scalar mass to zero, there is no symmetry protecting it from acquiring large (quadratically divergent) corrections at the loop level. These quadratic divergences make the low energy model with fundamental scalars very sensitive to the ultraviolet structure of the theory, and the small scalar masses are then unnatural. The solution of this dilemma can be either to abandon the concept of fundamental scalars, as in technicolor models, or to search for a theory where the offending quadratic divergences miraculously cancel. Since fermion loops have opposite sign as bosonic loops, a theory associating to each fermion a bosonic partner, and relating their couplings so as to ensure the cancellation of the quadratically divergent loop corrections associated to them, would be able to do the miracle. This is what is called supersymmetry, which is a symmetry relating bosons with fermions, and in so doing allows in some sense for chiral symmetry to protect also bosonic masses. An important point is that in order that the quadratically ‘divergent’ loop corrections to the SM Higgs boson mass do not exceed the electroweak scale (and hence the expected value of the Higgs boson mass), i.e. $(g^2/16\pi^2)\Lambda^2 < (250 \text{ GeV})^2$, the cutoff signaling the energy at which supersymmetry should enter to play a role should be $\Lambda_{SUSY} < (4\pi/g)250 \text{ GeV} \sim \text{TeV}$. Hence, the solution of the naturalness problem requires that supersymmetry be present at the weak scale, and not much above.

---

5Even in the case that we mentioned before in which the doublet Higgs boson is kept light in a GUT theory by making it a Goldstone boson of a spontaneously broken global symmetry of the scalar potential, this symmetry does not prevent the appearance of quadratically divergent corrections to its mass from diagrams involving gauge or Yukawa interactions.
3.2. Supersymmetric multiplets

What supersymmetry does is to pair together fermionic and bosonic fields, and mix them up through the supersymmetric transformations, which can schematically be written as

\[ \delta_{SUSY} \text{Fermion} = \text{Boson}, \quad \delta_{SUSY} \text{Boson} = \text{Fermion}. \]  

(38)

On the other hand two supersymmetric transformations can produce a boost since

\[ \{ \delta_{SUSY}, \delta_{SUSY} \} = \text{boost}, \]  

(39)

and hence supersymmetry is intimately related to Poincaré invariance. There is indeed a theorem (due to Haag, Lopuszański and Sohnius) stating that the most general symmetry of the \( S \) matrix is just \( \text{SUSY} + \text{Poincaré} + \text{internal symmetries} \). Moreover, the fact that supersymmetric transformations can generate boosts indicates that making supersymmetry a gauge symmetry (so that the transformation parameters depend on the space-time point) should lead to invariance under general coordinate transformations and hence automatically include General Relativity. Local supersymmetry (supergravity) may then be the road to incorporate gravity into the SM.

Since SUSY transforms fermions into bosons and vice-versa, it requires the existence of fermionic generators \( Q_\alpha \), and the above mentioned relation between two SUSY transformations and a boost results from the SUSY algebra relation \( \{ Q_\alpha, \tilde{Q}_\beta \} = 2 \sigma_\alpha^\mu P_\mu \) (with the dot over the index just indicating that it refers to the conjugate of the generator, and \( \sigma^\mu = (1, \vec{\sigma}) \)). A useful concept to write down supersymmetric lagrangians is that of superspace, introduced by Salam and Strathdee [14]. The superspace includes a fermionic ‘coordinate’ \( \theta_\alpha \) which is somehow conjugate to the supersymmetric generator \( Q_\alpha \), in the same sense that the usual coordinates \( x_\mu \) are conjugate variables to \( P_\mu \). The supersymmetric particle multiplets are then described by superfields defined in the superspace \((x_\mu, \theta_\alpha)\).

The basic multiplets of a supersymmetric theory are:

- The chiral supermultiplet \( \Phi(x, \theta) \) contains a complex scalar \( A \), a Weyl fermion \( \Psi \) and an auxiliary scalar \( F \) (this last is not propagating since its equation of motion is just algebraic, involving no derivatives, and hence it can be eliminated in favor of the other two propagating fields). Under an infinitesimal global supersymmetric transformation with parameter \( \xi \) one has, as anticipated in Eq. (38), that \( \delta A = \sqrt{2} \xi \Psi \) and \( \delta \Psi = \sqrt{2} (\xi F + i \sigma^\mu \xi \partial_\mu A) \). On the other hand, the transformation of the auxiliary field is a total derivative, \( \delta F = \sqrt{2} \partial_\mu \Psi \sigma^\mu \xi \).

- The real vector multiplet \( V^a \) is used to describe gauge fields \( A^a_\mu \), and has the associated fermionic partners \( \lambda^a_\mu \) known as gauginos, which will then be in the adjoint representation of the gauge group. The superfield associated to them has also an auxiliary component, the \( D^a \) scalar field, which also transforms into a total derivative under an infinitesimal SUSY transformation.

- The graviton supermultiplet involves the spin two graviton \( G^\mu_\nu \) paired to its superpartner, the spin 3/2 gravitino.

The trick to write down a supersymmetric Lagrangian is to exploit the fact that auxiliary \( F \) terms of chiral multiplets and \( D \) terms of vector multiplets transform under SUSY as total derivatives, and are hence good Lagrangian densities. For instance, one can construct a vector superfield from the following combination of the matter chiral superfields \( \Phi_i \) and gauge boson vector superfield \( V^a \),

\[ \sum_{i,j} \Phi_i^\dagger \exp[2g T^a_{ij} V^a] \Phi_j. \]  

(40)

It can be seen that the \( D \) term (auxiliary part) of this superfield, when written in terms of component fields, includes the usual fermion auxiliary covariant derivative involving the coupling to the gauge bosons, and is
hence the appropriate supersymmetric generalization of the gauge invariant fermion kinetic term. This Lagrangian density also includes new terms such as a fermion-sfermion-gaugino coupling (with the same strength as the previous one as required by supersymmetry). Furthermore, it also includes a quartic scalar coupling $\sum a(D^a)^2/2$, where

$$D^a = g \sum_i \Phi_i^a T^a \Phi_i.$$  \hfill (41)

The strength of this quartic scalar coupling is $g^2$, as required in order that the quadratic divergences in the sfermion self energies coming from the scalar loop associated to this quartic coupling (summed to the loop involving the gauge bosons) cancels with the one coming from the fermion-gaugino loop.

The gauge kinetic terms result from the $F$ component of the chiral field obtained from the square of the so-called field-strength chiral field, $W = \lambda + (D + i\sigma_\mu F_{\mu
u}/2)\theta/2 + \theta^2 \sigma^\mu \partial_\mu \lambda/4$.

Finally, the Yukawa couplings result from the $F$ term of the superpotential $W$, which is a generic cubic (so as to be renormalizable) polynomial in the chiral fields, i.e.

$$W(\Phi) = \frac{1}{2} m_{ij} \Phi_i \Phi_j + \frac{1}{3} \lambda_{ijk} \Phi_i \Phi_j \Phi_k.$$ \hfill (42)

In order for $W$ to be a chiral superfield, it is necessary that it be an holomorphic function of the chiral superfields $\Phi_i$, not involving the conjugate fields, and this has important phenomenological implications. For instance, in the minimal supersymmetric standard model (MSSM) two Higgs doublets of opposite hypercharge are required to give masses to up-type and down-type quarks. This is different than what happens in the SM, where one can freely use the conjugate Higgs doublet $\sigma_2 H^*$ (which has opposite hypercharge as $H$) together with $H$ for this purpose, but this is not derivable from a superpotential in a supersymmetric framework. Another reason justifying the need for two Higgs doublets is that the higgsino fermionic partners contribute to triangle anomalies, and hence two higgsinos with opposite charges are required to cancel them.

Besides giving rise to the Yukawa couplings $\mathcal{L}_Y = (1/2)(\partial^2 W/\partial \Phi_i \partial \Phi_j) \Psi^i \Psi^j$, the superpotential also generates a contribution to the scalar potential depending on $F_i = \partial W/\partial \Phi_i$. The complete scalar potential is then

$$V = \sum_i |F_i|^2 + \frac{1}{2} \sum_a (D^a)^2.$$ \hfill (43)

Again the quadratic divergences of the scalar fermion masses associated to the quartic couplings present in $|F|^2$, which are proportional to $\lambda^2$, cancel with those from the ordinary fermion loops induced by the Yukawa couplings $\lambda$.

### 3.3. Supersymmetry breaking

Following the procedure described in the previous section one can then write down a supersymmetric Lagrangian involving the standard model fields and their superpartners (the scalar fermions: sleptons, sneutrinos and squarks; the gauge fermions: gluinos, winos and binos; the Higgs fermions or higgsinos). Clearly an immediate drawback of this would be the appearance of all these states with the same charges and masses as the SM ones but different spin, which are not observed in nature. This implies that supersymmetry has to be broken, so as to lift the degeneracy in mass inside supermultiplets and make the superpartners sufficiently heavy and beyond present experimental bounds. However, to retain the good properties of SUSY, this breaking has to be ‘soft’, i.e. without reintroducing quadratic divergences [15]. This means that only some kind of SUSY violating terms in the Lagrangian are allowed, such as bilinear or trilinear scalar couplings (but not quartic ones) and gaugino masses (but not fermion Yukawa couplings without their associated $|F|^2$ terms in the potential). The SUSY soft breaking Lagrangian can
then be written as

$$-\mathcal{L}_{soft} = \sum_{\text{scalar}} m_i^2 |\phi_i|^2 + (AW^{(3)} + BW^{(2)} + \sum_a M_a \lambda_a \lambda_a + h.c.),$$  \hspace{1cm} (44)

where $M_a$ are gaugino masses, while

$$W^{(3)} = \lambda_u H_2 \tilde{Q} \tilde{u}^c + \lambda_d H_1 \tilde{Q} \tilde{d}^c + \lambda_t H_1 \tilde{L} \tilde{e}^c \hspace{1cm} (45)$$

contains the cubic terms in the superpotential, but with the superfields replaced by their scalar components. Similarly, the quadratic term is $W^{(2)} = -\mu H_1 H_2$.

We have not specified the family structure of the different couplings, but once this one is taken into account, it is found that the most general softly broken supersymmetric Lagrangian contains more than one hundred new unspecified parameters, clearly introducing a lot of arbitrariness in the model building. Moreover, generic parameters can induce large FCNC effects and violate bounds on $m_K = m_{\tilde{K}}$, on $K^0 \rightarrow \mu^+ \mu^-$ or $\mu \rightarrow e\gamma$, they can lead to large $CP$ violating electric dipole moments for the neutron or the electron, they can affect $K^0 \rightarrow \pi^0 \pi^0$, etc.. In particular, the absence of a generalized GIM mechanism to suppress FCNC can be understood from the fact that the fermion-fermion-neutral gauge boson coupling remains diagonal after a unitary rotation of the fermion fields, and the sfermion-sfermion-gauge boson also after a unitary rotation of the sfermion fields, but however the fermion-sfermion-gaugino vertex will not be diagonal in the flavor indices after going to the mass eigenstate basis.$^6$

The way to break spontaneously supersymmetry without breaking Lorentz invariance is to induce a non-vanishing VEV for the auxiliary component of a superfield, i.e. either $\langle F \rangle \neq 0$ ($F$ breaking) or $\langle D \rangle \neq 0$ ($D$ breaking). In this case the ground state energy will be non-vanishing (see Eq. (43)), what is the signal of global supersymmetry breaking.

In global supersymmetry breaking there is an important constraint on the resulting mass spectrum, which is known as the tree-level supertrace formula, i.e.

$$\text{STr} \mathcal{M}^2 \equiv \sum_{\text{bos}} \mathcal{M}^2 - \sum_{\text{ferm}} \mathcal{M}^2 = 0.$$  \hspace{1cm} (46)

For instance, considering just a chiral multiplet with a scalar component $\varphi = S + iP$, the typical spectrum after supersymmetry breaking will be to have the scalar and pseudoscalar components splitted with respect to the fermionic particle in such a way that $M_{S,P}^2 = M_{F}^2 \pm \Delta M^2$, i.e. $M_S^2 + M_P^2 - 2M_F^2 = 0$ (the two in the fermionic contribution to the supertrace comes from the two d.o.f. of a Weyl fermion). This clearly cannot be implemented in the observable sector of the supersymmetric SM, since it would imply for instance the existence of scalar fermions lighter than the corresponding fermions (or with negative squared masses). Hence, the usual strategy is to break supersymmetry in a hidden sector not directly coupled to the SM fields, and communicate it to the observable sector by gravitational effects (gravity mediation), at the loop level by means of gauge interactions (gauge mediation) or for instance exploiting the superconformal anomaly (anomaly mediation). A cartoon of this general framework is shown in Fig. 10, which was drawn after a week of rain during the meeting at the Itacuruça Island: the hidden sector is represented by the sky above the clouds, which was totally disconnected from the observable sector. The thick layer of clouds represent the messenger fields, which are in contact with both sectors and once in a while let some photons go through, allowing the breaking of susy (the sunshine) to be transmitted into the observable sector, giving rise to the tenuous light that we could see.

$^6$Alternatively, one may diagonalize the fermion-sfermion-neutral gaugino vertices and then be constrained to work with off-diagonal sfermion mass matrices.
Fig. 10: SUSY breaking in hidden sector and its mediation to the observable sector.
The simplest way to suppress FCNC effects is by having a mechanism of supersymmetry breaking mediation giving rise to universal scalar masses $\tilde{m}^2$, so that in first approximation the mass eigenstates can be freely rotated. In gravity mediation scenarios, this is usually attributed to the universal character of gravitational interactions, that should then generate a common soft mass for all scalar fermions (also generation independent $A$ and $B$ soft terms are predicted). In gauge mediation scenarios, the scalar masses arise at two-loops and turn out to be proportional to the gauge coupling constant ($m^2_f \propto \alpha_f^2$), so that for instance squarks are degenerate among themselves as a result of the blindness of gauge interactions to the family indices.

It has to be noticed however that the universal conditions will hold at the messenger scale (i.e. at the Planck scale for gravity mediation, or at a smaller scale $M \geq 10^2$ TeV for gauge mediation), but these masses run for decreasing scales and in so doing induce non-universalities in the mass spectrum. One then generally predicts anyway the presence of non-zero FCNC effects at low energies, which in gravity mediation schemes are sometimes even at the verge of contradicting experimental bounds. Regarding the gaugino masses, one often assumes for simplicity that there is a unification relation, so that they become all equal at the GUT scale ($M_3 = M_2 = M_1$) in gravity mediation scenarios, while in gauge mediation ones they arise at one loop and are hence proportional to the corresponding gauge couplings.

We see then that the new parameters of a supersymmetric model can be significantly restricted in specific models. For instance, in the minimal supersymmetric standard model (MSSM) with gravity mediation one has just the five parameters $\tilde{m}^2, M_\lambda, A, B$, and $\mu$.

A lot of work has been devoted to try to understand the origin of soft terms and to relate the low energy predictions with the possible mechanisms of supersymmetry breaking and mediation, which are the biggest unknowns for supersymmetry model builders.

### 3.4. The supersymmetric Higgs sector

As we already mentioned, the supersymmetric SM requires the existence of at least two complex Higgs doublets with opposite hypercharges

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^+ \end{pmatrix}, \quad H_2 = \begin{pmatrix} H_2^0 \\ H_2^+ \end{pmatrix}. \quad (47)$$

This represents 8 d.o.f., and when the electroweak symmetry gets broken 3 will be eaten by the massive gauge bosons $W^\pm$ and $Z$, leaving one physical charged Higgs $H^\pm$ and three neutral ones (one pseudoscalar $A$ and two scalars, $h$ and $H$).

The neutral Higgs potential will have the following contributions:

- The soft mass terms: $m_{H_1}^2 |H_1^0|^2 + m_{H_2}^2 |H_2^0|^2$
- The bilinear soft term: $BW^{(2)} = -B \mu H_1^0 H_2^0$
- The $|F|^2$ terms

$$\sum_i |F_i|^2 \geq \left| \frac{\partial W^{(2)}}{\partial H_1^0} \right|^2 + \left| \frac{\partial W^{(2)}}{\partial H_2^0} \right|^2 = \mu^2 \left( |H_1^0|^2 + |H_2^0|^2 \right). \quad (48)$$

- The $D$ terms

$$\frac{1}{2}D^2 = \frac{1}{2} \left( g^2 \left( T_3(H_1^0) |H_1^0|^2 + T_3(H_2^0) |H_2^0|^2 \right)^2 + g^2 \left( Y(H_1^0) |H_1^0|^2 + Y(H_2^0) |H_2^0|^2 \right)^2 \right). \quad (49)$$
This leads to the scalar potential

\[ V(H_1^0, H_2^0) = m_1^2 |H_1^0|^2 + m_2^2 |H_2^0|^2 - m_3^2 (H_1^0 H_2^0 + h.c. ) + \frac{g^2 + g'^2}{8} \left( |H_1^0|^2 - |H_2^0|^2 \right)^2, \]  

(50)

with \( m_{1,2}^2 \equiv m_{H_{1,2}}^2 + \mu^2 \) and \( m_3^2 \equiv B \mu \). Notice that the direction \( |H_1^0| = |H_2^0| \) is ‘D flat’, and in order that the potential be bounded from below along it one needs to satisfy \( m_1^2 + m_2^2 > m_3^2 \). On the other hand, the origin is unstable (and hence the electroweak symmetry is broken) as long as \( m_1^2 m_2^2 < m_3^2 \).

At the GUT scale these two conditions are incompatible under the assumption of universality, i.e. if \( m_1^2 = m_2^2 = \tilde{m}^2 + \mu^2 \), but a quite remarkable property of the MSSM is that in the running to low energies, the effects of the large top Yukawa coupling pushes down the parameter \( m_3^2 \), which can become negative leading to what is known as the radiative breaking of the electroweak symmetry. Actually this desirable feature provided the first theoretical hint for the need to have a large top mass [16], at a time when most phenomenologist believed that the top mass had to be at the few tens of GeV level.

The minimization of the Higgs potential leads to a vacuum state with

\[ \langle H_1^0 \rangle = \frac{v}{\sqrt{2}} \cos \beta , \quad \langle H_2^0 \rangle = \frac{v}{\sqrt{2}} \sin \beta, \]  

(51)

with

\[ \sin^2 \beta = \frac{2 m_3^2}{m_1^2 + m_2^2} \]  

(52)

and in order to reproduce the correct electroweak scale one needs to satisfy

\[ \frac{M_Z^2}{2} = g^2 + g'^2 v^2 = m_1^2 - m_3^2 \tan^2 \beta \frac{\tan^2 \beta}{\tan^2 \beta - 1}. \]  

(53)

Equation (52) is usually employed to trade the parameter \( B \) by \( \tan \beta \equiv \langle H_2^0 \rangle / \langle H_1^0 \rangle \), while the constraint in Eq. (53) is usually employed to express \( \mu^2 \) in terms of the remaining susy parameters (leaving \( \text{sign}(\mu) \) undetermined). Hence, after imposing the electroweak symmetry breaking constraints the additional parameters present in the gravity mediated MSSM, assuming universality and unification, are

\[ \tilde{m}^2, \quad M_\lambda, \quad A, \quad \tan \beta, \quad \text{sign}(\mu). \]  

The mass spectrum in the Higgs sector is directly obtained by expanding the scalar potential around its minimum, and in so doing one finds that the lightest Higgs boson mass satisfies \( m_h < M_Z |\cos 2\beta| \). This tree-level relation (which is essentially excluded by present LEP bounds) is sizeably affected by one-loop corrections to the scalar potential arising from the large top Yukawa coupling. These lead to a contribution \( \delta m_h^2 \simeq (3/\pi^2)(m_t^4/v^2)\log(m_t/v) \), which can bring the Higgs mass above present bounds, but anyhow a generic prediction of the MSSM is that the lightest Higgs should be lighter than \( \sim 130 \) GeV. Hence, the experimental search for a Higgs boson in this mass range is a very important test for supersymmetry, since failing to find it would exclude the most natural models of weak scale supersymmetry.

Another important aspect of the radiative corrections to the Higgs potential is that they reduce its overall scale-dependence, since as we mentioned before the tree level parameters \( m_t^2 \) where running significantly (some even changing sign) as the weak scale was approached, and hence the Higgs spectrum and couplings obtained from the tree level potential would have a strong scale dependence, but this is cured by the radiative corrections.
The Higgs boson searches at LEP have focused mainly on the Higgs-strahlung process \((e^+e^- \rightarrow Z \rightarrow Zh)\), from which essentially all the kinematically allowed range \(m_h < \sqrt{s} - M_Z \approx 113\) GeV has been excluded (for \(\tan\beta < 8\), since otherwise the \(ZZh\) coupling is suppressed). For large \(\tan\beta\) the preferred discovery channel is \(Z \rightarrow hA\), i.e. the light Higgs production in association with the pseudoscalar \(A\), and this has excluded the range \(m_h < 90\) GeV.

### 3.5. The supersymmetric particle spectrum

To obtain the sparticle spectrum one has to run all soft breaking parameters from high energies (assuming some proper boundary conditions, probably motivated by the absence of FCNC, simplicity and predictability) down to the weak scale. A plausible spectrum arising from this exercise for gravity mediated scenarios is depicted in Fig. 11. The lightest superpartner is the lightest of the four Majorana neutralinos (i.e. the mass eigenstates which are mixtures of the four neutral fermions which are the partners of the photon, the \(Z\) and the two neutral complex Higgses). The charged fermions which are partners of the charged Higgses and of the \(W\) mix into two charginos, which are Dirac fermions since they are charged. Colored particles get splitted from uncolored ones due to the running associated to strong interactions, and in this way gluinos are typically much heavier than charginos or neutralinos, and squarks are much heavier than sleptons. Third generation squarks are lighter than those of the first two generations due to the effects of the Yukawa couplings.

Fig. 11: Plausible particle spectrum in a gravity mediated supersymmetric model.

Since this doubling of degrees of freedom associated to the superpartners is expected to appear below the TeV scale, upcoming colliders (Tevatron run II, LHC, \(e^+e^-\) linear colliders) are ideal to search for these particles, which indeed constitute one of their main targets.
3.6. SUSY and gauge unification

One of the most appealing hints in favor of supersymmetry at the weak scale is related to the effects that it has on the running of the gauge couplings. In a supersymmetric scenario, the experimentally observed gauge couplings do indeed unify at a unique scale, and this one is large enough to be consistent with proton decay constraints (see Fig. 12).

![Diagram of running of the gauge couplings in the SUSY SM at one loop, adopting $T_S = 1$ TeV.](image)

This can be seen from the change in the $\beta$ functions arising from the additional contributions produced by loops involving gauginos and sfermions, which for instance make SU(3) somewhat less asymptotically free ($b_3 = -3$ instead of the non-susy value of $-7$), and make the SU(2) beta function positive ($b_2 = 1$), while for $U(1)_Y$ one has $b_1 = 33/5$, as can be easily checked using Eq. (22). These changes delay the unification of the couplings up to $M_{GUT} \approx 10^{16}$ GeV, and work provided the threshold scale $T_S$ above which the supersymmetric spectrum appears is light enough, $T_S < \text{TeV}$.

3.7. SUSY and proton decay

We saw that in non-supersymmetric GUTs proton decay was due to four fermion operators of the form $qqql$ which resulted from the contraction of the propagator of the very heavy triplet Higgs or gauge bosons. Being these operators of dimension $\dim = 6$, they were suppressed by two powers of the GUT scale (i.e. $\propto M_{GUT}^{-2}$). These lead to $\tau(p \to e^+\pi^0) \sim 10^{31} \text{yr}(M_{GUT}/10^{15}\text{GeV})^4$, which was in contradiction with observations for the non-supersymmetric GUT scale of $10^{14}$ GeV. In SUSY, the GUT scale becomes $10^{16}$ GeV, and hence there is no conflict with proton decay arising from $\dim = 6$ operators. In the supersymmetric case there are however new operators, involving two bosons (squarks or sleptons) and two fermions, which can induce proton decay, and since these are $\dim = 5$, they are suppressed by only one power of the GUT scale (i.e. by the contraction of one heavy fermionic propagator, such as the one from higgsino triplet exchange). Hence, $\dim = 5$ operators turn out to be the dangerous ones. Since initial and final states involve ordinary quark and leptons, the $\dim = 5$ operators have to be ‘dressed’ with light susy partners, as is shown in Fig. 13. This typically results in the proton decay rate being proportional to $(m_\tilde{g} \tan \beta / \tilde{m}^2)^2$, and hence it is quite sensitive to the supersymmetric spectrum (increasing for heavier gluinos!).
Another important fact is that writing the operators in terms of superfields, which obey Bose statistics, one can realise the need to antisymmetrize them with respect to the family indices. For instance, the operator \( O_1 = e^{ijk}(u_i^a d_j^\beta - d_j^\alpha u_i^\beta)(u_k^c \ell - d_k^\gamma \nu) \) is acceptable since it is antisymmetric in both color \((i, j, k)\) and flavor \((\alpha, \beta, \gamma)\) indices. This implies that the dominant proton decay channel in supersymmetric models is not into pions (involving first generation quarks) but into kaons. Typical predictions for the lifetime of this channel are

\[
\tau(p \to K\bar{\nu}) \sim 10^{29} \div 10^{35} \text{ yrs.}
\]  

The Superkamiokande bound of \( \tau(p \to K\bar{\nu}) > 6.7 \times 10^{32} \) yrs then excludes part of the supersymmetric parameter space, disfavoring in particular heavy gluinos for large \( \tan \beta \), with the bounds depending on the assumed scale for the scalar masses \( \tilde{m} \).

### 3.8. R parity symmetry, collider searches and dark matter

One important difficulty that has to be faced when trying to write down a supersymmetric version of the standard model is that there are Yukawa-like couplings which are consistent with gauge and super symmetries but violate baryon or lepton number. The allowed trilinear couplings are

\[
\mathcal{L}_R = \lambda_{ijk} Q_i L_j U_k^c + \lambda'_{ijk} U_i^c D_j^c D_k^c + \lambda''_{ijk} L_i L_j E_k^c.
\]  

Hence, unlike what happens in the SM, in its supersymmetric version \( B \) and \( L \) are not automatic symmetries of the Lagrangian. If the above mentioned couplings are allowed, they can give rise to extremely fast proton decay, mediated by the exchange of squarks or sleptons with weak scale masses, so that very strong constraints result on the product of some pairs of couplings (typically \( \lambda \cdot \lambda < 10^{-27} \)). To make these couplings so small is unnatural, so that the simplest solution is to eliminate them directly by invoking a symmetry, the so-called R-parity. R-parity distinguishes between standard model particles, which are R-even (e.g. \( f \leftrightarrow f \) and \( H \leftrightarrow H \) under R parity), and their superpartners, which are R-odd (e.g. \( \tilde{h} \leftrightarrow -\tilde{h} \) or \( \tilde{f} \leftrightarrow -\tilde{f} \)). Hence, imposing R symmetry the couplings in Eq. (56) are forbidden, while those of the standard Yukawa couplings involving the Higgs doublet are allowed.

The discrete R-parity symmetry need not necessarily be imposed by hand, and may be the left over, after suy breaking, of a continuous R symmetry (i.e. a symmetry of the susy Lagrangian which also transforms the Grassman variables \( \theta \)) or it may be related to an underlying gauge symmetry. For instance, in SO(10) models the allowed Yukawa couplings involve the \( 16 \times 16 \times 10 \), and this does not include the \( R \) violating couplings (that would be in the \( 16 \times 16 \times 16 \), which is not SO(10) invariant).

The conservation of R-parity has some very important phenomenological implications:

- It requires that superpartners be always produced in pairs, and hence this increases the threshold for their production.
- It implies that in the decay of a superpartner there should be always a superpartner (or an odd number of them). This requires in particular that the lightest superpartner (LSP) has to be stable. The stability of the LSP, which in supergravity models is usually the lightest neutralino and in gauge mediation models can be the gravitino, is crucial for the experimental searches, since it leads to the
characteristic missing energy signature, associated to the escape from the detector of the weakly interacting neutral LSP. For instance, typical signatures of supersymmetry at $e^+e^-$ colliders involve the production of pairs of charginos $\chi^+ \chi^-$ or sleptons $\tilde{e}^+ \tilde{e}^-$, with their subsequent decay into leptons or jets plus neutralinos. The first ones are observed but the neutralinos (and eventually some neutrinos) are not, and their effect is to take away significant amounts of energy, which will be missing in the overall budget. Strongly interacting particles such as squarks and gluinos are best searched at hadronic colliders, through the processes $q\bar{q}, gg, gg \rightarrow \tilde{q}\tilde{q}, \tilde{g}\tilde{g}, \tilde{q}\chi, \tilde{g}\chi$, with subsequent cascade decays of the superpartners. Another ‘background free’ process is the trilepton signal with missing energy associated to the chargino/neutralino pair production, i.e. $p\bar{p} \rightarrow \chi^{\pm}_1 \chi^0_2$, with $\chi^\pm \rightarrow \ell^\pm \nu \chi^0_1$ and $\chi^0_2 \rightarrow \chi^0_1 \ell^+ \ell^-$. 

Up to the present no signal of superpartner production has been observed, so that roughly speaking charginos and sleptons have to be heavier than $\sim 100$ GeV, while squarks and gluinos must be above $\sim 200 \div 300$ GeV. A major improvement in the mass reach will be achieved with the LHC, allowing to test superpartner masses up to the TeV scale, so that it may not be unreasonable to say that supersymmetry at the weak scale (as required by the naturalness problem) will have to be discovered (or discarded) by this machine.

- Another major implication of the LSP stability is its possible relevance to explain the dark matter which is known to be present in the Universe. Within supersymmetric models, the dark matter can naturally be attributed to LSPs which are left over from the early stages of the hot Universe. At that time all particle species where initially in thermal equilibrium and the weakly interacting ones were ‘frozen out’ little after they became non-relativistic and in this way they were able to survive up to the present times. The exact amount of relic particles that remains depends essentially on the rate of annihilations of neutralinos since this determines how the decoupling takes place. Larger annihilation rates clearly imply less surviving particles, and for a weakly interacting massive particle (WIMP, the neutralino being the preferred candidate), one can show that the contribution to the relic density is $\Omega \simeq 10^{-37}$ cm$^3$/s/GeV, but of course several subtleties enter into a detailed computation (many possible channels, threshold effects, resonant annihilations, coannihilations among different superpartners, etc.).

At any rate, it is remarkable that weak scale supersymmetry, which has its motivations in theoretical issues completely unrelated with the dark matter problem, predicts the existence (at least in R parity conserving scenarios) of a natural candidate for cold dark matter. On the other hand, if neutralinos indeed constitute the galactic dark matter, this offers a new possibility to experimentally search for supersymmetry trying to detect the WIMPs and to characterize their properties. Several groups are trying to observe directly the nuclear recoils resulting from halo WIMPs interacting inside low background detectors (using germanium, sodium iodine, superconducting granules, liquid xenon, etc.). Others (such as Superkamiokande or MACRO) are trying to observe the high energy neutrinos produced in the annihilations of the dark matter neutralinos which have become trapped by the Sun or the Earth in the last five Gyrs. The predicted rates depend strongly on the supersymmetric parameters, but for instance may be in the range of the sensitivity of upcoming direct detection experiments for large values of $\tan\beta$, especially for positive $\mu$, in which case the spin independent coherent contribution to the scattering cross sections is enhanced. There is even a claim of a positive signal having been observed by the DAMA Collaboration, but this is partly in conflict with results from other groups.

### 3.9. Supersymmetry in virtual processes

Besides producing superpartners at colliders or trying to detect those produced in the big bang, supersymmetry may manifest through the effects that virtual superpartners may have on some particular processes. Two important cases are the flavor changing transition $b \rightarrow s\gamma$ and the anomalous magnetic moment of the muon. The first one takes place already at the loop level in the SM, involving in particular the exchange of the heavy top quark, and the supersymmetric contri-
bution can naturally be of the same order of magnitude. The diagrams contributing to it are shown in Fig. 14. The loop involving the charged Higgs has the same sign as the SM contribution, the

\[ \text{Fig. 14: SM and supersymmetric contributions to } b \to s\gamma. \text{ The photon can be attached to any charged particle line.} \]

one involving neutralino or gluino exchange is usually negligible while the one involving chargino exchange grows with $\tan \beta$, and hence is the leading one for large $\tan \beta$. It interferes destructively with the SM piece for $\mu > 0$ and constructively for $\mu < 0$. Since the measured rates are already somewhat below the SM expectations (see Eq. (8)), one may say that large $\tan \beta$ with negative $\mu$ is in conflict with observations, and that positive $\mu$ is somewhat preferred.

\[ \text{Fig. 15: Supersymmetric contributions to the muon anomalous magnetic moment.} \]

Regarding the anomalous magnetic moment of the muon, the supersymmetric contribution arises from the diagrams in Fig. 15. Since these are flavor and CP conserving processes they are hard to avoid, and they are of a similar size as the electroweak contributions. The recent Brookhaven results on $a_\mu$ where the first to be sensitive at the level of the electroweak corrections, and the 2.6 standard deviation discrepancy with respect to the SM expectations they reported then clearly attracted a lot of attention from the supersymmetric community. The magnetic moment has associated with it a flip in the chirality of the fermion, and hence this is why the anomalous magnetic moment of the electron, although measured more accurately, is not a good place to look for supersymmetry ($a_{e}^{SUSY} \sim (m_e/m_\mu)^2 a_{\mu}^{SUSY}$). The left-right transition in $a_{\mu}^{SUSY}$ can take place also through the higgsino vertex (proportional to $\lambda_\mu \sim m_\mu \tan \beta$) or through a L–R smuon mass insertion in the scalar propagator (proportional to $m_\mu^{2}_{LR} = m_\mu (A - \mu \tan \beta)$). Hence the SUSY contribution is relevant for large $\tan \beta$, with typical predictions being $|a_\mu(SUSY)| \leq 14 \times 10^{-10}(100 \text{GeV}/\tilde{m})^2 \tan \beta$. The reported results (after correcting with the new theoretical estimates [17]), $a_\mu(BNL) - a_\mu(SM) = 25(16) \times 10^{-10}$, then suggest that $\tan \beta$ should be large and that the supersymmetric mass scale $\tilde{m}$ should be below a few hundred GeV, with two superpartners (one of them being charged) being then at the reach of future collider experiments. The supersymmetric explanation of $a_\mu(BNL)$ also prefers positive values of $\mu$, and this goes in the same direction as $b \to s\gamma$ results, and as we saw is also good news for future experiments looking for supersymmetric dark matter.

\[ \text{^7New theoretical estimates have changed a sign in the contribution involving photon photon scattering, reducing this discrepancy to less than two standard deviations [17].} \]

111
In conclusion, the SM has been extremely successful in accounting for existing experimental data, and has been subject to several precision tests. With the energy frontier being pushed above the electroweak scale by upcoming experiments, and rare processes becoming tested with high sensitivity, we are entering the stage in which the search for new physics is becoming the main experimental goal in particle physics. Certainly some revolution in our understanding of the underlying laws of nature will come out from this.

ACKNOWLEDGEMENTS

I would like to thank the students and the conference organizers for having made possible this very interesting school. Many thanks also to Gian Giudice for comments on the manuscript. This work was partially supported by Fundación Antorchas.

BIBLIOGRAPHY

The present ‘light’ notes on the basic aspects of some of the major attempts to extend the SM have the aim to just motivate the (experimental and theoretical) students to undertake the study of this very broad subject, about which many books and specialized review articles have been written. Here below are listed a few of them which can be useful for this purpose, and in them also a comprehensive list of references to the original works can be found:

References

[1] Most of the experimental results quoted are taken from the Review of Particle Physics, 


T. Yanagida. in Proc. Unified Theory and Baryon number in the Universe (KEK 1979).


INTRODUCTION TO NEUTRINO PHYSICS

Paolo Lipari
Dipartimento di Fisica, Università di Roma “la Sapienza”, and I.N.F.N., Sezione di Roma, P.A. Moro 2, I-00185 Roma, Italy

Abstract
Neutrino Physics in recent times has been going through a revolutionary period. Measurements of neutrinos coming from the sun or produced by cosmic rays in the Earth’s atmosphere have revealed that neutrinos “oscillate”, that is they change their flavor periodically with time. The observation of this phenomenon allows to obtain precious information on the neutrino masses and their mixing, shedding information about the “flavor problem”, that is the origin of the fermion families and of the masses and mixing of quarks and leptons. In these lectures we introduce the subject and review these recent developments.

1. INTRODUCTION
In many ways neutrinos are special particles, first of all they are several order of magnitude lighter than all other fermions, and in fact no direct measurement has found evidence for a non zero neutrino mass: for all detected neutrinos one has measured (within errors) $p c = E$. Neutrinos are also neutral, do not feel the strong interactions, and interact only weakly, in fact extraordinarily weakly. For example the neutrinos produced in nuclear reactors with energy $E_\nu \sim 1$ MeV have a cross section of order $\sigma \sim 10^{-44}$ cm$^2$, this corresponds to a probability $\sim 10^{-18}$ to interact in a solid detector with the thickness of one meter, or a probability $\sim 10^{-11}$ to interact inside the Earth traveling along a trajectory that passes through its center.

Neutrinos are then “elusive” particles, but they also are very abundant particles in the universe, and we do not notice them only because they interact so little. We live immersed in a bath of cosmological neutrinos produced when the universe was hot and young, actually ‘immersed’ is probably not a good expression because the neutrino densities inside and outside our bodies are essentially identical. The average number density of cosmological neutrinos is $n_\nu \simeq 336$ cm$^{-3}$, and inside our galaxy it can be much larger because of gravitational clustering effects (that are determined by the $\nu$ mass and are not known). Approximately 60 billions of solar neutrinos, produced in the core of the sun by nuclear reactions, cross each square centimeter of our body every second. A truly fantastic number of neutrinos ($\sim 10^{58}$) is emitted in few seconds following the gravitational collapse of a massive star that triggers one supernova explosion. The sun and the closest supernova in the last three centuries (SN1987A in the Large Magellanic Cloud) have been the first two astrophysical objects to be “seen” in neutrinos, and in the (hopefully) not too far future, new detectors should be able to detect neutrinos from other astrophysical sources, and neutrinos will become a new precious “messenger” from distant objects. The smallness of the neutrino interaction cross section is at the same time the biggest problem for the newly born science of neutrino astronomy, since very massive detectors are required to have appreciable event rates, and also an extraordinary opportunity, because neutrinos can emerge from deep inside the core of astrophysical objects, revealing directly the physical processes that operate there. For example solar neutrinos come directly from the sun inner core, and directly tell us about the nuclear fusion reactions that are the source of the solar luminosity, while the visible photons are emitted from the the surface with a black body spectrum.

Neutrinos have two type of interactions, they can couple with a $Z^0$ boson, changing their 4-momentum but keeping their identity (neutral current interactions), or they can couple with a $W^\pm$ boson “transforming” into one of the charged leptons $e^\pm$, $\mu^\pm$ or $\tau^\pm$ (charged current interactions). Of course
the interaction vertices can generate different type of processes. For example the vertex of \( W \) boson, a neutrino and a charged lepton results in processes such as: a charged lepton transforming into a neutrino with the emission of a (virtual) \( W \) boson, or a \( W \) boson that generates a neutrino charged lepton pair. Because of the charged current processes one can define the concept of neutrino flavor. The flavor of a neutrino is simply the type (that is the mass) of the charged lepton that is connected to the same charged current vertex. For example in the decay of a (virtual) \( W^- \) or \( W^+ \) one conventionally writes:

\[
W^+ \rightarrow \ell^+ + \nu_\ell, \quad W^- \rightarrow \ell^- + \bar{\nu}_\ell,
\]

where the label \( \ell = e, \mu, \tau \) of the neutrino correspond to the mass (or type) of the charged lepton. The distinction of \( \nu \) and \( \bar{\nu} \), and the \( \ell \) label given to the neutrino has a very simple and clear experimental justification. It has been found that a “\( \nu_e \)”, defined according to the convention of (1), interacting close to its creation point can only produce \( e^- \) (and not a \( \mu^- \) or an \( e^+ \)), while a “\( \bar{\nu}_e \)” can only produce \( e^+ \), a \( \nu_\mu \) only a \( \mu^- \), and similarly for the other neutrino types. For example a crucial experiment demonstrated that the neutrinos produced in a pair with a \( \mu^+ \), close to their creation point can only generate \( \mu^- \) in the sequence of processes:

\[
\pi^+ \rightarrow \mu^+ + \nu_\mu, \\
\nu_\mu + p \rightarrow \mu^- + \text{hadrons}.
\]

In summary, both the flavor label and the “bar” notation have a well defined phenomenological meaning.

Neutrinos have half integer intrinsic angular momentum (or spin). The extraordinary discovery of the 50’s has been the finding that all “\( \nu \)” have their spin, within experimental uncertainties, anti-parallel to their momentum, while for all “\( \bar{\nu} \)” the spin is parallel. In different words all, \( \nu \) are left-handed, and all \( \bar{\nu} \) are right-handed. This is at the basis of the “chiral nature” of the weak interactions, and is the source of the effects of violation of parity observed in nature. This fact opens a fascinating possibility, namely that the states \( \nu \) and \( \bar{\nu} \) are the two different spin states of a two state “Majorana particle” that is the anti-particle of itself, while all the other fermions are “Dirac particles” having 4 states (particle–antiparticle, each with two independent spin states).

A remarkable physical effect, originally predicted by Bruno Pontecorvo [1] has been recently measured in neutrino physics, and is the existence of “flavor oscillations”. To understand this phenomenon let us consider for a moment the decay of a (virtual) \( W \) boson into a quark pair. It is kown that this decay is of type:

\[
W^- \rightarrow \bar{d} + d', \quad W^- \rightarrow \bar{s} + s', \quad W^- \rightarrow \bar{b} + b',
\]

where for example the \( d' \) quark that is produced in association with a \( \bar{d} \) is a linear superposition of the three quark states that have charge \(-1/3\) and well defined mass: \( d, s \) and \( b \) (and similarly for \( s' \) and \( b' \)). The states produced at the \( W \) vertices are connected to the mass eigenstates by the Cabibbo–Kobayashi–Maskawa matrix [2]:

\[
\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V^{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.
\]

It is natural to expect that this “mixing” phenomenon can also happen in the lepton sector, and the neutrinos \( \nu_\ell \) produced in association with the charged lepton \( \ell^- \) will not have a well defined mass but be a linear combination of the states that have a well defined mass:

\[
\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U^{\text{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix},
\]

where \( U^{\text{PMNS}} \) is the Pontecorvo–Maki–Nakagawa–Sakata matrix [1,3], the leptonic analogous of the
CKM matrix for quarks. In the case of neutrinos, because of the smallness of the neutrino masses a new remarkable quantum-mechanical phenomenon can exist, that is the phenomenon of flavor oscillations. In the case of quarks, the measurement of the final state can easily determine the mass of the quark, on the other hand it is practically impossible to measure the mass of a neutrino, therefore the amplitudes for the production of neutrinos with different mass must be added coherently. Since the amplitudes of different mass components evolve differently with space and/or time, acquiring different quantum mechanical phases, it follows that flavor is a periodical function of time. In other words if a neutrino of type $\nu_e$ is produced at a point $A$, a measurement close point to this creation point will result in the production of charged leptons of type $e^-$, but at a more distant point $B$ the very same neutrino can generate also $\mu^-$ or $\tau^-$ particles if the quantum mechanical phases of its different components have become significantly different from each other. The closer the neutrino masses are to each other, the longer is the time needed to have appreciable phase differences between the different components, and (since neutrinos travel at approximately the speed of light) the probability for a transition to a different flavor becomes significantly different from zero only for a long distance between the production and detection points. On the other hand, experiments with long pathlength allow the measurement of tiny mass differences between the neutrinos.

The remarkable development of the last few years has been the discovery of the phenomenon of flavor oscillations with two type of experiments detecting “atmospheric” and solar neutrinos. The results are consistent with each other, and can be interpreted to give us very constraining information about the neutrino mass splittings and mixing. Additional controversial indications of flavor transitions of type $\nu_\mu \to \nu_e$ have been obtained by the LSND experiment.

These lectures present an introduction to the physics of neutrinos, with particular attention to the discovery of the flavor transitions. The work is organized as follows: in the next Section we discuss general properties of neutrinos, Section 3 gives a rapid overview of the sources of neutrinos in nature, Section 4 discusses the phenomenon of flavor oscillations in vacuum, Section 5 discusses in more detail the quantum mechanics of oscillations, Section 6 discusses neutrino oscillations in matter, Sections 7 and 8 discuss the observations of atmospheric and solar neutrinos, Section 9 and 10 additional observations of reactor and accelerator neutrinos, Section 11 discusses double beta decay, Section 12 presents a phenomenological summary on flavor transitions, Section 13 very briefly discusses model for the neutrino masses.

The bibliography on neutrino physics is very rich. Several good textbooks exist on the subject, for example [4–6]. Recent lectures/reviews of the subjects are: [7–13]. In order to follow the developments of the field two very useful sites on the web are [14] and [15].

2. GENERAL PROPERTIES OF NEUTRINOS

2.1. Neutrino Flavors:

In the standard model of particle physics there are three neutrinos flavors. Each neutrino completes a doublet with the corresponding charged leptons. For example a $W^+$ gauge boson can couple to the three neutrino/charged-lepton pairs:

$$
W^+ \rightarrow e^+\nu_e, \\
\rightarrow \mu^+\nu_\mu, \\
\rightarrow \tau^+\nu_\tau.
$$

(6)

The label associated to the neutrinos correspond to the flavor, that is the mass of the charged lepton produced in association.
Electron neutrinos

Historically the first evidence of the neutrinos appeared in the study of nuclear beta decay in the 1920’s. Beta decay is a form of radioactivity where a primary nucleus \((A, Z)\) decays into a lighter secondary one of the same mass number but charge differing by one unit, and an electron (or positron) and nothing else visible:

\[
(A, Z) \rightarrow (A, Z \pm 1) + e^\mp + \text{nothing else visible}. \tag{7}
\]

An example is this process is the decay of a free neutron into a proton and an electron with a lifetime of approximately 15 minutes (886.7 sec). In the absence of additional invisible particles in the final state, the \(e^\pm\) should have a well defined energy \(E_e \approx Q \equiv M_i - M_f\), on the contrary experimentalists found a continuum spectrum ranging from \(m_e\) up to the maximum allowed energy \(Q\). Various explanations were offered for this anomaly (Niels Bohr even arrived to speculate about the possibility that energy conservation was violated) until in december 1930 Wolfgang Pauli suggested that a non-observed particle was emitted in the reaction. This unobserved particle had to be neutral (to conserve electric charge), very light (since \(E_{e,\text{max}} \approx Q\)) and with spin 1/2 (to satisfy angular momentum conservation and statistics). In modern notation beta decays are now considered as the processes:

\[
\begin{align*}
n &\rightarrow p + e^- + \bar{\nu}_e, \\
A(Z, N) &\rightarrow A(Z + 1, N - 1) + e^- + \bar{\nu}_e, \\
A(Z, N) &\rightarrow A(Z - 1, N + 1) + e^+ + \nu_e. \tag{8}
\end{align*}
\]

Soon after the proposal of Pauli, Enrico Fermi [16] proposed a model of beta decay (and weak interactions in general) in terms of the product of two weak “currents”, one connecting the initial and final state nucleons, and the other one the final state electron and anti-neutrino. The strength of the interaction was related to the coupling strength \(G_F\). In Fermi’s theory the two currents connected at the same space-time point, assuming that the interaction was very short ranged. In modern language the Fermi constant \(G_F\) can be understood as the effect of the exchange of a virtual but massive heavy “photon”, that is the gauge boson \(W^\pm\) with a mass \(M_W \approx 80\) GeV. The weakness of the weak interaction can be understood as the effect of the large mass (and corresponding short range) of the \(W\) boson with \(G_F \approx e^2/M_W^2\), or more exactly:

\[
\frac{G_F}{\sqrt{2}} = \frac{1}{8} \left( \frac{e}{\sin \theta_W} \right)^2 \frac{1}{M_W^2}, \tag{9}
\]

(where \(\theta_W\) is the Weinberg angle). Fermi’s theory could successfully describe the observed spectra of electrons and positrons emitted in beta decays, and at the same time it could predict the existence of other processes involving the “weak currents”, in particular the existence of the “inverse \(\beta\) decay” processes such as:

\[
\begin{align*}
\bar{\nu}_e + p &\rightarrow e^+ + n, \\
\nu_e + n &\rightarrow e^- + p. \tag{10}
\end{align*}
\]

The cross section for these processes could be predicted and found to be extremely small. The detection of neutrinos therefore remained elusive for more than 20 years, until the results of Reines and Cowan [17] in 1956. They used the nuclear reactor at Savannah River in South Carolina as an anti-neutrino source and a detector of water with dissolved cadmium chloride; the detection technique was the coincident (with a few \(\mu\)sec delay) observation of a positron emitted in the reaction \(\bar{\nu}_e + n \rightarrow e^+ + n\), together with a photon emitted in the deexcitation of cadmium after the neutron capture: \(n + ^{112}\text{Cd} \rightarrow ^{113}\text{Cd} + \gamma\). In 1996 Fred Reines received the Nobel prize in physics for this discovery. The study of \(\beta\) decays allowed to establish the “chiral nature” of neutrinos, the \(V–A\) structure of weak interactions, and the existence of Parity non-conservation effects in the weak interactions.
**Muon neutrinos**

Muon neutrinos and anti-neutrinos are produced pion and muon decay:

\[
\pi^+ \to \mu^+ + \nu_\mu, \\
\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu.
\]  

(11) \hspace{1cm} (12)

When the decay \( \pi^\pm \to \mu^\pm + \nu \) was discovered it was natural to ask the question if the undetected particles produced in the decay were identical with the neutrinos produced in \( \beta \) decay. This problem was solved experimentally in 1962 by L. Lederman, M. Schwarz and J. Steinberger [18] who performed the first experiment with accelerator neutrinos. Bombarding a Beryllium target with a 15 GeV primary proton beam, they produced a beam of pions, that because of the decay was the source of a neutrino beam, and then studied the interaction with nucleons of type \( \nu_\mu + N \to \mu^- + X \) or \( \nu_\mu + N \to e^- + X \). Only the first type of interactions were found, demonstrating that \( \nu_e \) and \( \nu_\mu \) are different particles. The reason why the decay \( \pi^+ \to e^+ \nu_e \) that has a larger phase space available is strongly suppressed (with a branching ratio \( \sim 10^{-4} \)) is connected to the \( V - A \) nature of the weak interaction, see later in Section 2.4.

**Tau neutrinos**

The \( \nu_\tau \) is associated to the third charged lepton, the \( \tau^- \) discovered in 1975 at Stanford. Tau neutrinos and anti-neutrinos are expected to be produced in \( \tau^- \) decays.

\[
\tau^- \to \nu_\tau + W^-, \\
\to \nu_\tau + e^- + \bar{\nu}_e, \\
\to \nu_\tau + \mu^- + \bar{\nu}_\mu, \\
\to \nu_\tau + [\pi^0].
\]  

(13)

**2.2. Number of light neutrino species**

How many families of quarks and leptons exist in Nature? The number of light neutrino species that have the usual electroweak interactions can be determined studying the process:

\[ Z^0 \to \nu_\alpha + \bar{\nu}_\alpha. \]  

(14)

The partial width of this decay \( \Gamma_{\nu\bar{\nu}} \) is accurately calculable in the standard model:

\[ \Gamma_{\nu\bar{\nu}} = 166.9 \text{ MeV} \]  

(15)

and therefore the \( Z \) will have a branching ratio into invisible final states that is proportional to the number of light neutrino species:

\[ \Gamma_{\text{invisible}} = N_\nu \Gamma_{\nu\bar{\nu}}. \]  

(16)

The invisible width can be obtained experimentally subtracting the contributions of all visible channels from the measured total width:

\[ \Gamma_{\text{invisible}} = \Gamma_{\text{tot}} - \Gamma_{\text{vis}} = 498 \pm 4.2 \text{ MeV}. \]  

(17)

This can be interpreted [19] as the result that the number of the light, active \( \nu \) species is:

\[ N_\nu = \frac{\Gamma_{\text{inv}}}{\Gamma_{\nu\bar{\nu}}} = 2.994 \pm 0.012. \]  

(18)
2.3. Sterile neutrinos

Equation (18) gives the number of light particles that have the standard properties of neutrinos with respect to the weak interactions, and therefore does not apply to “sterile neutrinos”. A sterile neutrino is a particle that is a singlet with respect to the $SU(3) \otimes SU(2) \otimes U(1)$ group of the standard model, but that is mixed or related to neutrinos. As an example (among many) of sterile neutrinos one can consider the “Exact Parity Model” [21], where the gauge group of the universe has the form:

$$G = G_L \otimes G_R^R = [SU(3) \otimes SU(2) \otimes U(1)]_1 \otimes [SU(3) \otimes SU(2) \otimes U(1)]_2,$$

(19)

and each fermion has a mirror partner, so that the the ordinary fermions are singlet with respect to the subgroup $G_R^R$, while the mirror particles are singlet with respect to $G_L$. Because of the existence of the mirror particles the model is symmetric with respect to parity. In this model there are three “mirror neutrinos”, that do not interact with the ordinary $W$ and $Z$ bosons and are coupled only to the mirror gauge bosons $\bar{W}$ and $\bar{Z}$. Clearly the bound (18) is not relevant for these particles. It can be shown that in this model the eigenstates of the Hamiltonian are linear combinations of type

$$\{1, 2\} = \{\nu \pm \bar{\nu}\}/\sqrt{2}$$

of standard and sterile states, and therefore one expects for all all three flavors oscillations $\nu_\ell \leftrightarrow \bar{\nu}_\ell$ (with a mixing angle $\approx 45^\circ$).

An (indirect) limit on the existence of sterile neutrinos mixed with ordinary ones comes from cosmology, and in particular from Big–Bang nucleosynthesis. From the measured primordial abundances of helium-4 and other light nuclear species, one can obtain an estimate of the energy density of the universe at the epoch of nucleosynthesis, that can be translated in an estimate of $N_{\nu}$ that is equivalent number of light ($m < 1$ MeV) species that are in thermal equilibrium at this epoch. For a larger $N_{\nu}$ the evolution of the universe becomes faster and the helium abundance grows. From the observed primordial helium abundance one can then extract an upper limit:

$$N_{\nu} < 3.3,$$

(20)

but perhaps more conservatively $4 \nu$ species can still be tolerated). In the absence of mixing, sterile neutrinos that interact only extremely weakly will not be in thermal equilibrium with the cosmic soup at the nucleosynthesis epoch, and therefore the limit (20) cannot in general exclude their existence, however oscillations between standard and sterile neutrinos can generate a significant population of sterile neutrinos. If the mixing is sufficiently large and the oscillations sufficiently fast (that is if the mass differences are sufficiently large) one can violate the bound (20). Therefore the bound (20) be used to set limits on the existence of oscillations between standard and sterile neutrinos.

2.4. Chirality and Helicity

It has been observed that all neutrinos have (within errors) spin parallel to their momentum, while for anti-neutrinos spin and momentum are antiparallel. This is a consequence of the $V - A$ structure of the weak interaction. In this theory the spinor of the fermions (anti-fermions) produced in a $W^\pm$ vertex are projected with $\{1 \mp \gamma_5\}$ ($\gamma_5$ is a Dirac matrix). In words, particles are produced in weak interaction vertices with a well defined chirality. For a particle with $m/E$ small, a state of left chirality (that is the projection $\{1 - \gamma_5\}$) corresponds to an amplitude $\sim 1$ of having the spin antiparallel to the momentum and an amplitude $\sim m/E$ of having the opposite spin state. For right chirality the situation is reversed. Clearly in the case of massless particles chirality (that is the $\{1 \pm \gamma_5\}$ projections of a fermion state), and helicity (that is the projection of the spin parallel to the fermion momentum) coincide, however in general the two concepts are different, and this difference can have very important consequences.
An example one can consider the decay of the charged pion. This decay has two possible modes:

\[
\pi^+ \rightarrow \mu^+ + \nu_\mu ,
\]
\[
\pi^+ \rightarrow e^+ + \nu_e .
\]

The strong suppression of the second mode, is an outstanding manifestation of the \( V - A \) nature of the charged current weak interactions, and a clear illustration of the difference between the chirality and the helicity. Let us consider (see Fig. 1) the decay of pions at rest: \( \pi^+ \rightarrow \ell^+ + \nu_\ell \). The \( V - A \) structure of weak interactions requires the emitted \( \nu_\ell \) to be of left handed chirality. For \( m_\nu \simeq 0 \) this also means that it has the left handed (or negative) helicity (spin antiparallel to its momentum). Conservation of total angular momentum then requires \( \ell^+ \) to be of negative helicity. However, the \( \ell^+ \) is an antiparticle, and again for the \( V - A \) structure of weak interactions it must be produced in a state of right-handed chirality. Therefore the amplitude of the process must be proportional to the admixture of left handed (negative) helicity for a right-handed chirality charged-lepton, that is proportional to its mass: \( A(\pi^+ \rightarrow \ell^+ + \nu_\ell) \propto m_\ell \). Including phase space effects \( \propto (m_\pi^2 - m_\ell^2) \) one has the expectation:

\[
R_\pi \equiv \frac{\Gamma(\pi^+ \rightarrow e^+ + \nu_e)}{\Gamma(\pi^+ \rightarrow \mu^+ + \nu_\mu)} = \left( \frac{m_e}{m_\mu} \right)^2 \left( \frac{m_e^2 - m_\mu^2}{m_\pi^2 - m_\mu^2} \right)^2 = 1.28 \times 10^{-4} .
\]

In agreement (after including a 4% radiative correction) with the experimental value: \( R_\pi = (1.230 \pm 0.004) \times 10^{-4} \).

\[ \text{2.5. Limits on the Neutrino masses} \]

\[ \text{Direct Measurements} \]

A direct measurement of the neutrino masses is in principle possible, with kinematical methods, determining the momentum and energy of a neutrino. These methods have only produced upper limits for the neutrino masses [19].

For \( \nu_e \) (or more precisely \( \overline{\nu}_e \)) the most sensitive method is the study of the shape of the high energy part of the \( \beta \) decay spectrum. The “end point” of the spectrum (the highest kinematically allowed energy for the electron) depends on the \( \nu \) mass \( (E_{e_{\text{max}}} \simeq Q - m_\nu) \). The most studied decay (because of a low \( Q \) value) is the decay of Tritium:

\[
^3\text{H} \rightarrow ^3\text{He} + e^- + \overline{\nu}_e .
\]

The current limits\(^1\) are

\[
m_{\nu_1} < 2.5 \text{ eV at } 95\% \text{ c.l. (Troitsk)} ,
\]
\[
m_{\nu_2} < 6 \text{ eV at } 95\% \text{ c.l. (Mainz)} ,
\]
\[
m_{\nu_3} < 15 \text{ eV at } 95\% \text{ c.l. (PDG suggestion)} .
\]

\(^1\nu_1, \nu_2 \text{ and } \nu_3 \text{ are the primary mass components of } \nu_e, \nu_\mu \text{ and } \nu_\tau, \text{ respectively.}\)
A limit on the mass of $\nu_\mu$ can be obtained studying the decay $\pi^+ \to \mu^+ \nu_\mu$. In the pion rest frame the momentum of the muon is:

$$E_\mu^* = \frac{m_{\pi^+}^2 + m_\mu^2 - m_\nu^2}{2 m_{\pi^+}}. \quad (24)$$

The neutrino mass can be obtained from measurements of the masses $m_\pi, m_\mu$ and the momentum $p_\mu$ of the final state muon in the pion rest frame. This results in an upper limit:

$$m_{\nu_\mu} < 170 \text{ keV at } 90\% \text{ c.l.} \quad (25)$$

For $\nu_\tau$ the limit is obtained measuring the missing 4-momentum is decays like $\tau^- \to 5\pi^+ \nu_\tau$ where all particles except the neutrino are measures. This results in the limit:

$$m_{\nu_\tau} < 18.2 \text{ MeV at } 95\% \text{ c.l.} \quad (26)$$

**Neutrinoless double beta decay**

The non observation of the neutrinoless double beta decay allows to set a limit on the combination:

$$\langle m_{\nu_e} \rangle_{\text{eff}} = \sum_j U_{ej}^2 m_j \lesssim 0.4 \text{ eV}, \quad (27)$$

(for more discussion see Section 11.). This limits applies only to the case where neutrinos are Majorana particles (see Section 2.6). Note also that the limit depends on the neutrino mixing matrix and since the $U_{ej}$ are complex numbers cancellations are possible and $\langle m_{\nu_e} \rangle_{\text{eff}}$ is a lower limit for the highest neutrino mass.

**Supernova SN1987A**

The measurements of a small number of $\overline{\nu}_e$ from supernova 1987A in February 1987 has allowed to obtain a limit $m_{\overline{\nu}_e} \lesssim 20 \text{ eV}$ (see Section 3.2).

**Cosmological limit**

The most stringent limits on the neutrino masses can be obtained from cosmology (for more discussion see Section 3.1). The energy density in neutrinos can be very reliably estimated as:

$$\Omega_\nu = \frac{\rho_\nu}{\rho_c} = \frac{1}{h^2} \frac{\sum_j m_j}{93 \text{ eV}}, \quad (28)$$

where $\rho_c$ is the critical density that makes the universe flat, and $h \simeq 0.65$ is the Hubble constant in units of 100 Km/s/Mpc. The measured flatness of the universe gives a limit $\Omega_{\text{tot}} h^2 \lesssim 0.4$ that results in:

$$\sum_j m_j \lesssim 37 \text{ eV.} \quad (29)$$

A more stringent (but more model dependent) limit can be obtained from studies of structure formation in the universe. From these studies emerges a broad consensus that most of the energy density of the universe is in form of “dark energy” ($\Omega_\Lambda \sim 0.7$), and “cold dark matter” ($\Omega_{CDM} \simeq 0.25$), leaving only a smaller amount available for neutrinos. Recent measurements of the anisotropies of the Cosmic Microwave Background Radiation by WMAP [20] have been interpreted as giving a limit on the contribution of Hot Dark Matter in the universe:

$$\Omega_\nu h^2 \leq 0.0076 \quad (95\% \text{ C.L.}). \quad (30)$$

This corresponds to the limit $\sum_j m_j \leq 0.7 \text{ eV}.$
2.6. DIRAC or MAJORANA particles?

Can neutrinos be their own antiparticles? If this is the case they are “Majorana” particles, if not they are “Dirac” particles like the quarks and the charged leptons. It is known experimentally that neutrinos emitted in $\beta^-$ decay (which we conventionally call $\bar{\nu}_e$) are not captured in reactions which are caused by $\nu_e$. That is the sequence of reactions:

$$n \rightarrow p + e^- + \bar{\nu}_e$$
$$\bar{\nu}_e + ^{37}\text{Cl} \rightarrow ^{37}\text{Ar} + e^-$$

is not observed even if the reaction $\nu_e + ^{37}\text{Cl} \rightarrow ^{37}\text{Ar} + e^-$, does exist in nature. This does not imply that $\nu_e$ and $\bar{\nu}_e$ are different objects, because the weak interactions that are responsible for neutrino interactions are chiral (or have a $V-A$ structure). The particles which we call $\bar{\nu}_e$ are right-handed, while the particles that are absorbed by $^{37}\text{Cl}$ are left-handed, and therefore the non observation described in (31) could simply be due to a strong dynamical suppression (of order $\sim (m_\nu/E)^2$), even if $\nu_e$ and $\bar{\nu}_e$ are in fact the same object. The sequence of reactions described in (31) is therefore possible if two conditions are met:

- $\nu_e$ and $\bar{\nu}_e$ are identical particles (that is the neutrino is a Majorana particle)
- The neutrino mass is non vanishing, so that the helicity suppression is large but not complete.

Something like the sequence of reactions in (31) could be observed in the process of neutrinoless double beta decay (see Section 11.), and the study of this process appears the best way to verify the nature (Dirac or Majorana) of the neutrino.

3. NEUTRINO SOURCES

Neutrinos are very common particles in the universe, and play a very important role in many objects and events in astrophysics and cosmology. Figure 2 shows a plot of the energy spectrum of neutrinos that reach the surface of the Earth. The spectrum extends for more than 20 orders of magnitude in both energy and intensity. “Naturally” produced neutrinos have several components: particle accelerators.

3.1. Cosmological Neutrinos

The relic neutrinos, i.e. the neutrinos left over from the early epochs of the evolution of the universe, have a number density of about 56 cm$^{-3}$ for each neutrino species ($\nu_e$, $\bar{\nu}_e$, $\nu_\mu$, $\bar{\nu}_\mu$, $\nu_\tau$, $\bar{\nu}_\tau$) and a black-body spectrum with a temperature $T_\nu \simeq 1.947$ Kelvins (or an average kinetic energy of $5 \times 10^{-4}$ eV). [For a general review of cosmology see for example [22,23].]

The spectrum of relic photons from the big bang has been well studied by the COBE satellite and other detectors and has a temperature $T_{\gamma}^{\text{now}} = 2.728$ Kelvin. This corresponds to a number density:

$$n_\gamma = \frac{2}{(2\pi\hbar)^3} \int d^3p_\gamma \frac{1}{e^{E/T} - 1} \simeq 412 \text{ cm}^{-3}. \quad (32)$$

An analogous neutrino spectrum must exist from the early universe. In fact, when the universe was sufficiently hot, neutrinos where kept in equilibrium with the photons with reactions such as $\gamma \gamma \rightarrow \nu_\alpha \bar{\nu}_\alpha$. Neutrinos dropped out of thermal equilibrium when the temperature dropped below $T \sim 10^{10}$ Kelvin, and from this point remained essentially decoupled from the rest of the universe, but proceeded to cool as the universe expanded. The neutrino temperature is cooler than the photon one, because the photon component was “reheated” by the annihilation of the electron and positron components of the ‘cosmic
Fig. 2: Flux of neutrinos at the surface of the Earth. The three arrows show the energy thresholds for charged current interactions on a free proton target. The line that refers to cosmological neutrinos assumes that the neutrino mass is vanishing. For massive neutrinos the flux is modified since gravitational clustering enhances the density, the $\nu$ velocity is decreased and the energy spectrum is modified. The line that refers to Supernova neutrinos describes $\bar{\nu}_e$. Different neutrino species have similar spectra, with differences difficult to appreciate in the figure. The line that describes geophysical neutrinos includes the $^{238}\text{U}$ and $^{232}\text{Th}$ decay chains, the flux weakly depend on geopgraphical location. The atmospheric neutrino fluxes are calculated for the Kamioka location. Only the lowest energy part depends on the location. A range of prediction for the flux of astrophysical neutrinos is shown.
soup’ that happened at a temperature $T_\gamma \simeq m_e$. Since the entropy of the photon gas after the $e^\pm$ annihilation is equal to the sum of the entropy of the three components ($\gamma, e^+, e^-$) before annihilation, one has:

$$T_\gamma^3 \times 2 = T_\nu^3 \times \left[2 + 4 \times \frac{7}{8}\right]$$

or

$$T_\nu = \left(\frac{4}{11}\right)\frac{1}{3} T_\gamma.$$  \hspace{1cm} (33)

This corresponds to a neutrino number density:

$$n_\nu = \frac{1}{(2\pi h)^3} \int d^3p \frac{1}{e^{E/T} + 1} \simeq 56\text{ cm}^{-3}.$$  \hspace{1cm} (35)

Note that in the last equation we have considered a single spin state for the neutrino. If the neutrino is massive and is a Dirac particle, also an additional neutrino (and anti-neutrino) spin state exists, however since these ‘wrong helicity’ states are singlet with respect to the standard model, they interact extremely weakly, and are present in negligible number in the cosmic soup, so that the neutrino number density of neutrinos does not depend on whether they are Dirac or Majorana particles. For completeness we also observe that our derivation of the neutrino number density has assumed that $n_\nu \simeq n_\sigma$. More in general, it is possible to construct models where the universe has a large net lepton number and $(n_\nu - n_\sigma)/(n_\nu + n_\sigma)$ is large. In these cases the neutrino number density is modified.

The cosmological neutrino energy density depends critically on the $\nu$ masses. For $m_\nu \gg T_\nu^{\text{now}} \simeq 1.6 \times 10^{-4}$ eV, the kinetic energy of the neutrinos is negligible and the $\nu$ energy density is simply $\rho_\nu = m_\nu n_\nu$. In units of the critical density $\rho_c = \frac{3}{8\pi G} H^2$ (the density that makes the universe flat) one has:

$$\Omega_\nu = \frac{\rho_\nu}{\rho_c} = \frac{1}{h^2} \frac{\sum_j m_j}{93\text{ eV}},$$  \hspace{1cm} (36)

where $h \simeq 0.65$ is the Hubble constant in units of $100$ Km/sec/Mpc. Measurements of the age of the universe and of its expansion rate give a limit $\Omega h^2 < 0.4$ that sets the most stringent limit on the neutrino masses. On the other hand equation (36) tell us that if the neutrinos masses are of order of few eV’s or more they can be an important, or even the dominant component of the dynamics of the universe solving the so called “Dark Matter Problem”, one of the most important puzzles in Physics today.

If neutrinos are an important component of the mass of the universe they also play a crucial role in the formation of its observed structures (galaxies, clusters of galaxies, and larger scale coherent structures). At early times the universe was extremely smooth as demonstrated by the tiny amplitude of the temperature fluctuations of the cosmic microwave background radiation across the sky, gravity however enhances the density contrast of the irregularities in the density field, generating the observed structures. The outcome of this evolution depends on the initial spectrum of density fluctuations, but also on the composition of the cosmic fluid. If the matter that composes the cosmic fluid can move around easily, it can wipe out the fluctuations of the smaller scales. Neutrinos that are weakly interacting and that, because of their small mass, remain relativistic until later times are very efficient at erasing the initial fluctuations up to large scale, while other forms of dark matter, for example the WIMPS (Weakly Interacting Massive Particles) that move more slowly because of their larger mass are less efficient.

The recent measurements of the anisotropies of the Cosmic Microwave Background Radiation by WMAP [20] have been interpreted as giving a limit on the contribution of Hot Dark Matter in the universe: $\Omega_\nu h^2 \leq 0.0076$ (at 95% C.L.). This corresponds to the limit $\sum_j m_j \leq 0.7$ eV.

\footnote{The entropy per unit volume of a bosonic degree of freedom is $s_B = (2\pi^2/45) T^3$, and for a fermionic degree of freedom is $s_F = (7/8) s_B$.}
Limits on sterile neutrinos

Big bang nucleosynthesis [25] in one of the pillars of cosmology. According to the theory four elements: deuterium, helium-3, helium-4 and lithium-7 were formed when the universe was hot \((T \sim 1 \text{ MeV})\) and young \((t \lesssim 100 \text{ seconds})\). The calculated abundances of these elements, depend on the baryon density of the universe, and to reproduce the observed abundances one needs to conclude that baryonic matter is only a small fraction of the present energy density of the universe. This calculation allows also to put a limit on the “effective number” of light species (or “neutrinos”) in the universe. The fraction of the baryonic mass that is transformed into helium (that is observed to be approximately 25\%) grows with the number of effective neutrino species. This can be understood qualitatively observing that the evolution of the universe is speeded up if more particles species are present and the density is therefore larger, in a faster expansion the neutrons in the cosmic soup have less time to decay and end up forming more helium. Comparing the calculation with the observed helium abundance one obtains the result:

\[
N_{\nu}^{\text{eff}} \lesssim 3.3.
\]  

As discussed before, the measurements of the width of the Z boson at LEP set a very stringent limit on the number of neutrinos \(N_{\nu} = 2.994 \pm 0.012\), however the cosmological limit is more general because it applies to all light particle that are in thermal equilibrium in the universe, independently from their interaction properties, and therefore it applies to standard (or active) neutrinos, but also to any other light particles, including for example “sterile neutrinos”.

Does then the bound (37) rule out the existence of light sterile neutrinos, such as those implied by the LSND results? It is not necessarily so. The limit refer to a particle in thermal equilibrium with the soup of photons electrons and positron at the epoch of nucleosynthesis. A light sterile neutrinos that is interacting “super-weakly”, can only be put in thermal equilibrium with the rest of the matter by oscillations with ordinary neutrinos. Therefore the bound is effective only if the mixing of the sterile neutrino with an ordinary one is sufficiently large and if the squared mass difference is sufficiently large (so that oscillations have time to develop). Oscillations in the dense medium of the universe are controled by subtle matter effects whose calculation is still controversial, it is therefore still not clear what is the precise range of mixing parameters that can be excluded with the cosmological bound (37), and if they are in contrast with explanations in terms of sterile neutrinos of the LSND result\(^3\).

A very interesting topic that we cannot cover here is the role that neutrinos could play in baryogenesis. The observed excess of baryons over antibaryons in the universe may be related to decays of heavy Majorana neutrinos. For a review of the physics of cosmological neutrinos see [28].

### 3.2. Supernovae

Type II Supernovae explosions (for reviews and references see [29]) mark the end of the life of massive stars (with \(M \gtrsim M_\odot\)) that have developed an iron core surrounded by several onion-like burning shells and an outer envelope of hydrogen and helium. Iron is the most strongly bound nucleus in nature, and no additional burning can generate energy to support the star core. A stellar iron core is supported by the electron degeneracy pressure, a quantum mechanical effect related to the Heisenberg uncertainty principle: \(\Delta p \Delta x \simeq h\), that gives momentum to fermions squeezed in a small volume. When enough ‘nuclear ash’ has accumulated, and the iron core of the star reaches the Chandrasekhar limit of \(\sim 1.4\) solar masses, it becomes unstable, and collapses. The collapse is very rapid (can be approximated by free fall) and lasts only a small fraction of a second. The compressed core heats, ‘boiling’ the iron nuclei into separate nucleons, then it becomes energetically favorable to capture the electrons on free protons in the

\(^3\)More conservative analysis conclude that the bound of nucleosynthesis can be relaxed to \(N_{\nu}^{\text{eff}} < 4\), in this case one light sterile neutrino can certainly exist without conflict with cosmological data.
neutronization process
\[ e^- + p \rightarrow n + \bar{\nu}_e, \] (38)
that converts nearly all protons in the collapsing core into neutrons. The \( \nu_e \) produced in these reaction rapidly escape from the core generating a “neutronization burst” of \( \nu_e \) ’s. When the collapsing core reaches nuclear density (at a radius \( R \sim 10 \text{ Km} \)) the implosion is halted because of the stiffness of nuclear density matter. At this point a shock wave is formed that propagates outward ejecting the outer layers of the star and producing the spectacular visible explosion. The newly formed proto-neutron star has a radius \( R_{\text{n.s.}} \sim 10 \text{ Km} \) and therefore the density is of the same order of nuclear matter, and contains a kinetic energy of order

\[ E_{\text{kin}} = -E_{\text{grav}} = \frac{GM^2}{R_{\text{n.s.}}} \approx 3 \times 10^{53} \text{ erg}. \] (39)

Nearly all (99\%) of the energy is radiated away in the form of neutrinos, with only \( \sim 1\% \) going into producing the spectacular explosion as kinetic energy of the ejected layers and electromagnetic radiation. It is likely that the neutrinos emitted by the proto neutron star play a crucial role in the explosion, depositing enough energy near the outward propagating shock to “push” it out of the star, generating the explosion.

All six neutrino species contribute approximately equal to the energy outflow, since they are produced in the hot core by “flavor blind” processes like \( \gamma \gamma \rightarrow \nu_\alpha \bar{\nu}_\alpha \). The energy spectra are thermal, with average energies \( \langle E(\nu_e) \rangle \approx 11 \text{ MeV}, \langle E(\bar{\nu}_e) \rangle \approx 15 \text{ MeV}, \) and \( \langle E(\nu_\mu, \tau) \rangle \approx \langle E(\bar{\nu}_\mu, \tau) \rangle \approx 25 \text{ MeV}. \)

The neutrino emission lasts a time of order \( \Delta t \approx 10 \text{ seconds} \) that is determined by the time needed for neutrinos to “random walk” out of the core undergoing many scatterings in the dense material. The different average energies of the different components are the consequence of the different cross sections for elastic scattering with electrons (that still exist in the star). Electron neutrinos have the largest cross section, and are emitted from closer to the surface of the stars, while \( \nu_\mu, \tau \) ’s and \( \bar{\nu}_\mu, \tau \) with a smaller cross section are emitted from deeper inside the stars and are “hotter”.

The theory of neutrino emission in supernova explosions has had a dramatic confirmation the 23rd February 1987, when the neutrinos and the radiation of supernova (SN1987A) that had exploded 170,000 years before in the Large Magellanic Cloud (a small satellite galaxy of our Milky Way) reached the Earth (see Fig. 3). Two detectors: Kamiokande in Japan and IMB in the US detected a few events (11 Kamiokande, 7 IMB) in coincidence with each other and in a time interval of 13 seconds. These events that can be interpreted as the detections of positrons from the reaction \( \bar{\nu}_e + p = e^+ + n \). From the number and energy spectrum of the observed events, it is possible to extract (with large statistical errors) a fluence and a temperature for the \( \bar{\nu}_e \) emitted by the supernova with results in reasonable agreement with the theoretical predictions.

The handful of events detected from SN1987A have allowed to determine (or put limits) on several neutrino properties, and in particular on its mass. The velocity of a neutrino is given by:

\[ v = \frac{p_\nu}{\sqrt{p_\nu^2 + m_\nu^2}} \approx 1 - \frac{m_\nu^2}{2E_\nu^2}. \] (40)

Therefore two neutrinos of different energy emitted at the same time from a source at a distance \( L \) will reach the detector with a relative time delay:

\[ \Delta t = t_1 - t_2 = \frac{L}{v_1} - \frac{L}{v_2} \approx L \left( \frac{m_\nu^2}{2E_\nu^2} - \frac{m_\tau^2}{2E_\tau^2} \right). \] (41)

\(^4\)Most of these interactions are elastic scatterings with the neutrons in the star, where the energy of the neutrino changes, but little energy is transferred because of the neutron mass is much larger than the neutrino energy.
Fig. 3: Time and energy of the events detected the 23rd of February 1987 by the Kamiokande and IMB detectors that are associated with the explosion of the supernova SN1987A in the Large Magellanic cloud, approximately 170 thousand light years away.

Inserting appropriate numerical values one can see that for a mass of 10 eV, a 10 MeV neutrino emitted from SN1987A ($L \approx 55$ Kpc $\approx 1.7 \times 10^{21}$ cm) will have a delay of 2.8 seconds with respect to a high energy neutrino. An analysis of the data, shows no correlation of energy with time of arrival, and an overall duration of the signal in agreement with prediction, and therefore no indication of a non vanishing mass. From these considerations one can obtain a conservative limit:

$$m_{\nu_e} \lesssim 20 \text{ eV}. \quad (42)$$

3.3. Stars

Neutrinos are copiously produced in thermonuclear reactions which occur in the stellar interior and in particular in our sun. The reaction that liberates nuclear binding energy is the effective fusion reaction

$$4p + 2e^- \rightarrow ^4\text{He} + 2\nu_e + 26.73 \text{ MeV}.$$ 

So the sun luminosity implies a $\nu_e$ flux at the Earth of approximately $6 \times 10^{10}$ (cm$^2$ s)$^{-1}$. The detailed spectrum of the emitted neutrinos depends on the “path” taken by the nuclear reactions to burn hydrogen into helium. Most of the solar neutrinos have energy below 0.41 MeV, but a smaller component (of order $5 \times 10^6$ (cm$^2$ s)$^{-1}$) due to the beta decay of boron-8 extends up to 14 MeV, and plays a very important role in the detection of solar neutrinos. The measurements of solar neutrinos and their implications are discussed in Section 8.

3.4. Natural Radioactivity neutrinos

The Earth emits approximately 40 TeraWatt of energy, approximately 40% of this energy outflow is due to the decay of radioactive nuclei, 90% of this is due to decay chains due to the Uranium and Thorium decay chains. A $^{238}$U nucleus initiates a cascade of 8 $\alpha$ and 6 $\beta$ transitions that terminates in

---

5This limit corresponds to the assumption that electron neutrinos have a single dominant mass component.
the stable nucleus $^{206}$Pb, while $^{232}$Th initiates a cascade of 6 $\alpha$ and 4 $\beta$ transitions that terminates with $^{208}$Pb. In each $\beta$ decay is emitted a $\overline{\nu}_e$, with a maximum energy of 3.27 MeV. These decays induce at the surface of the Earth a flux of $\sim$ few $\times$ $10^6$ (cm$^2$ s)$^{-1}$. The flux will depend on the location of a detector, because we expect that chemical processes during the formation of the Earth have concentrated Uranium and Thorium in the Earth’s crust and (less strongly) mantle, depleting the core. Since the thickness of the crust is variable (being thicker under the continents and thinner under the oceans) the flux is dependent on detector location, for example being larger in Italy than in Japan. The high energy tail (above the threshold energy of 1.8 MeV) of this “geophysical” $\overline{\nu}_e$ flux can be in principle observed studying the capture on free protons followed by the detection of the neutron capture: ($\overline{\nu}_e + p \rightarrow e^+ + n$, $n + p \rightarrow \gamma + d$). New large mass detectors for the first time have a chance to measure the geophysical neutrinos. The KamLand detector in the Kamioka observatory in Japan with 1 Kton of CH$_2$ scintillator could measure a predicted rate of 50–100 events/year, while the Borexino detector in Italy with 300 tons of C$_9$H$_{12}$ scintillator could measure a rate of 10–20 events/year. These measurements can give information on the geophysical structure, dynamics and evolution of the present Earth [31].

### 3.5. Atmospheric Neutrinos

The Earth receives an approximately isotropic and constant flux of cosmic rays (with an intensity of approximately 0.5 particles/cm$^2$ sec sr). The primary cosmic rays, interacting in the upper atmosphere generate a number of secondary particles including charged pions and kaons that in turn generate neutrinos in decays of the type

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \mu^+ \rightarrow e^+ + \nu_e + \overline{\nu}_\mu$$

(and charged conjugate decays).

### 3.6. Astrophysical Neutrinos

Until today two astrophysical objects have been observed in neutrinos, the sun and the supernova SN1987A. In the future one expects the detection of other astrophysical sources of very high energy neutrinos. Very large (Km$^3$) neutrino telescopes [32,33] are currently under construction for the detection of high energy astrophysical neutrinos with $E_\nu \gtrsim 10^{12}$ eV (that is a Tera-electronVolt (TeV) and beyond). The motivation [32–33] for the existence of these very high energy neutrinos is the fact that fluxes of cosmic rays up to energy as high as $10^{20}$ eV are known to exist, and therefore must be produced somewhere, even if their sources have not been clearly established. Perhaps the best candidates as sources of the “Ultra High Energy” (UHE) cosmic rays are Active Galactic Nuclei or Gamma Ray Bursts. Our poor understanding of the origin of the cosmic rays is mostly due to the fact, that electrically charged particles are bent in the galactic ($|\vec{B}_{\text{gal}}| \simeq 3 \mu$Gauss) and inter-galactic magnetic fields and therefore do not point back to their sources. It is natural, in fact essentially unavoidable that the sources of UHE cosmic rays sources are also sources of neutrinos. This can be easily understood observing that in general a cosmic ray accelerator must contain a non negligible amount of “target material” in the form of gas, or also of photons in an environment filled with high densities of radiation. When an accelerated proton interacts with a target particle it will produce several charged pions (and kaons) that will then decay into neutrinos (with chains such as $\pi^+ \rightarrow \mu^+ \nu_\mu$ followed by $\mu^+ \rightarrow e^+ \nu_e \overline{\nu}_\mu$). The flux of astrophysical neutrinos has a harder energy spectrum than the flux of atmospheric ones, and one expects that it should become become dominant at energies above 1–10 TeV. Since we expect that most neutrino sources are extragalactic one expects a diffuse, approximately isotropic flux, due to the superposition of all sources in the universe, with the most powerful objects visible as neutrino point-like sources.

---

6In an astrophysical environment the density is in most case so low that even long lived particles can decay before reinteraction.
In order to detect these astrophysical neutrinos one needs very large and massive detectors. The most promising method is based on the detection of Cherenkov light in water or ice. Water (or ice) is naturally a beautiful candidate as a neutrino detector: it is dense, abundant and cheap, relativistic charged particles emit efficiently Cherenkov photons traveling in the medium, that is transparent in the relevant wavelength range (blue light). To construct a neutrino telescope it is therefore sufficient to distribute in a sufficiently large volume of water a number of photo-sensible devices (Photo Multipliers Tubes or PMT’s) capable to detect the Cherenkov photons of particles produced in neutrino interactions. The telescope must be screened from the background of the charged cosmic ray radiation, placing it deep under the Earth surface. The original concept of a very large water telescope for high energy neutrinos was developed by the DUMAND group, that originally planned to build a detector in the deep ocean (at a depth of ~ 4500 meters in the Hawaii archipelago) made of “strings” of photomultipliers. Each string should be anchored to ocean floor, and supported by buoyancy. Similar concepts are now pursued by three groups in the Mediterranean (the NESTOR (Greece), ANTARES (France), and NEMO (Italy) projects). The most advanced project (AMANDA) is located in Antarctica, at the South Pole and is based on the same Cherenkov method but uses ice instead of liquid water. Ice at sufficiently low temperature and high pressure, becomes transparent and is a good Cherenkov medium, and can be used also as the mechanical structure that supports the photomultipliers, that are placed in deep holes melted in the ice before its permanent refreezing.

3.7. Reactor and Accelerator neutrinos

Two important types of man-made sources of neutrinos also exist, namely: reactor neutrinos (see Section 9) and accelerator neutrinos (see Section 10).

4. NEUTRINO OSCILLATIONS IN VACUUM

The probability of finding a neutrino created in a given flavor state to be in the same state (or any other flavor state) can oscillate with time. The idea of the possible existence of neutrino oscillations was first introduced by Pontecorvo [1]. This remarkable phenomenon is the consequence of simple Quantum Mechanics. The neutrinos are produced in charged-current weak interactions in association with a charged lepton are weak-eigenstates: $\nu_e$, $\nu_\mu$ or $\nu_\tau$. In general these weak eigenstates do not have a well defined mass, and can be written as linear superpositions of three states $\nu_1$, $\nu_2$ and $\nu_3$ each having mass ($m_1$, $m_2$ and $m_3$). In general we can write:

$$|\nu_\alpha\rangle = \sum_j U_{\alpha j}^* |\nu_j\rangle ,$$

where we have indicated the flavor and mass eigenstates using respectively using greek and latin indices, and $U$, the neutrino mixing or Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, is a unitary matrix. The unitarity of $U$ follows from the fact that the three flavor (and mass) eigenstates are orthogonal to each other.

If the matrix $U$ is non trivial, it is simple to demonstrate that the phenomenon of flavor oscillation must exist. For simplicity let us assume that at time $t = 0$ a neutrino is produced as the flavor eigenstate $|\nu_\alpha\rangle$ with a well defined\footnote{The condition that the neutrino is produced with a well defined momentum is used here for pedagogical simplicity, and can be dropped. This is discussed together with additional details in Section 5.} 3-momentum $\vec{p}$ and consider the probability to find the neutrino in a different flavor state $|\nu_\beta\rangle$ at a later time $t$. The initial state at $t = 0$ can be written as a superposition of mass
eigenstates as:

\[ |\nu(0)\rangle = |\nu_\alpha\rangle = \sum_j U_{\alpha j}^* |\nu_j\rangle. \tag{45} \]

Note that the energy of our neutrino state is not well defined since each component will have an energy

\[ E_j = \sqrt{p^2 + m_j^2}. \]

The time evolution of a mass eigenstates (that has a well defined energy) is simply a phase factor \( e^{-iE_j t} \), and for our state created with well defined flavor the time evolution gives:

\[ |\nu(t)\rangle = \sum_j U_{\alpha j}^* e^{-iE_j t} |\nu_j\rangle. \tag{46} \]

Now the different components acquire different phases, and this results in non trivial effects. The probability amplitude of finding the neutrino at the time \( t \) in a flavor state \( |\nu_\beta\rangle \) is (we will now use the convention that repeated indices are summed over):

\[ A(\nu_\alpha \to \nu_\beta; t) = \langle \nu_\beta | \nu(t) \rangle = \langle \nu_\beta | \sum_k U_{\beta k} U_{k j}^* e^{-iE_j t} |\nu_j\rangle = U_{\beta j}^* U_{\alpha j} e^{-iE_j t}, \tag{47} \]

(where we have used \( \langle \nu_k | \nu_j \rangle = \delta_{jk} \)). The oscillation probability is obtained squaring the amplitude:

\[ P(\nu_\alpha \to \nu_\beta; t) = |A(\nu_\alpha \to \nu_\beta; t)|^2 = |U_{\beta j}^* U_{\alpha j} e^{-iE_j t}|^2, \tag{48} \]

is periodical in time and clearly is non vanishing even for \( \beta \neq \alpha \).

### 4.1. Two flavor case

It is instructive to consider in more detail the simpler case of the mixing of only two neutrino flavors (for example \( \nu_e, \nu_\mu \)). In this case we have only two mass eigenstates \((\nu_1, \nu_2)\) and two mass eigenvalues \((m_1 \text{ and } m_2)\). The relation between flavor and mass eigenstates can then be written in matrix form as:

\[ \begin{bmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{bmatrix}, \tag{49} \]

where we have used the compact notation \( c = \cos \theta, s = \sin \theta \), with \( \theta \) the mixing angle. Let us now assume now that an electron neutrino is created at time \( t = 0 \) with momentum \( p \). This corresponds to the initial state:

\[ |\nu(t = 0)\rangle = |\nu_e\rangle = c|\nu_1\rangle + s|\nu_2\rangle. \]

The two mass components of the neutrino have energies \( E_1 \) and \( E_2 \) given by:

\[ E_i = \sqrt{p^2 + m_i^2} \approx p + \frac{m_i^2}{2p} \approx E + \frac{m_i^2}{2E}, \tag{50} \]

After a time \( t \) the neutrino state will be:

\[ |\nu(t)\rangle = c e^{-iE_1 t} |\nu_1\rangle + s e^{-iE_2 t} |\nu_2\rangle. \tag{51} \]
The phase difference between the two components results in a non-trivial flavor evolution of the neutrino. For example, the probability of finding the neutrino with the muon flavor can be obtained as:

\[
P(\nu_e \rightarrow \nu_\mu; t) = |\langle \nu_\mu | \nu(t) \rangle|^2
\]

\[
= |\langle - iE_2 t - e^{-iE_1 t} \rangle|^2
\]

\[
= 2 c^2 s^2 \{ 1 - \cos[(E_2 - E_1)t] \}
\]

\[
= \sin^2 2\theta \sin^2 \left[ \frac{\Delta m^2}{4E} t \right],
\]

where we have used the notation \( \Delta m^2 = m_2^2 - m_1^2 \). For relativistic neutrinos one can also approximate \( L \approx t \). Another convenient form of the expression for the transition probability is

\[
P(\nu_e \rightarrow \nu_\mu; L) = \sin^2 2\theta \sin^2 \left[ 1.27 \Delta m^2 \frac{L}{E} \right],
\]

where \( L \) is in meters and \( E \) in MeV (or \( L \) is in km and \( E \) in GeV).

Examples of these 2-flavor oscillation probabilities are shown in Figs. 4 and 5. Note how the shape of the probability is defined by two parameters: the amplitude of the sinusoidal oscillations that is equal to \( \sin^2 2\theta \) where \( \theta \) is the mixing angle, and the oscillation length, that is the distance between any two closest minima or maxima of the oscillation probability. The oscillation length is inversely proportional to the squared mass difference \( \Delta m^2 \) and is linear with the neutrino energy \( E \):

\[
\ell_{\text{osc}} = \frac{4\pi E}{|\Delta m^2|} \approx 2.48 \frac{E \text{ (MeV)}}{|\Delta m^2| \text{ (eV}^2)} \text{ meters} = 2.48 \frac{E \text{ (GeV)}}{|\Delta m^2| \text{ (eV}^2)} \text{ Km}.
\]

Fig. 4: Plot of the oscillation probabilities as function of the neutrino pathlength \( L \) for the transitions \( \nu_\mu \rightarrow \nu_\mu \) and \( \nu_\mu \rightarrow \nu_\tau \) in the approximation of two flavor mixing. The energy of the neutrino is fixed.
Fig. 5: Example of the shape of the oscillation probabilities $P(\nu_\mu \to \nu_\mu)$ and $P(\nu_\mu \to \nu_\tau)$ as function of the neutrino energy $E_\nu$ for a fixed neutrino pathlength $L$. In the example we have chosen $\Delta m^2 = 1 \times 10^{-3}$ eV$^2$ (approximately one third the value indicated by the atmospheric neutrino experiments) and $L = 730$ Km (the $\nu$ pathlength of the Fermilab to Minos and Cern to Gran Sasso projects).

4.2. Dirac and Majorana Phases

In the case of three flavors, the neutrino mixing matrix $U$ can be parametrized in terms of 3 mixing angles, one CP violating phase and, if the neutrinos are Majorana particles, 2 additional “Majorana phases”. In the most general case ($n$ flavors) one has $n(n - 1)/2$ mixing angles, $(n - 1)(n - 2)/2$ “Dirac phases” and (possibly) $(n - 1)$ additional Majorana phases.

The most general $n \times n$ complex matrix depends on $2n^2$ real parameters; If the matrix is unitary $n^2$ parameters can be eliminated using the set of equations $U_{\alpha\beta}U^*_{\beta\alpha} = \delta_{\alpha\beta}$. The $n^2$ remaining parameters can be divided into $n(n - 1)/2$ angles and $n(n + 1)/2$ phases. If the neutrinos is a Dirac particle, $2n - 1$ phases can be removed with a proper rephasing of the left handed fields. This is easy to see, considering the set of equations that relate the flavor and mass eigenstates:

$$
|\nu_e\rangle = U^*_{e1}|\nu_1\rangle + U^*_{e2}|\nu_2\rangle + U^*_{e3}|\nu_3\rangle,
|\nu_\mu\rangle = U^*_{\mu1}|\nu_1\rangle + U^*_{\mu2}|\nu_2\rangle + U^*_{\mu3}|\nu_3\rangle,
|\nu_\tau\rangle = U^*_{\tau1}|\nu_1\rangle + U^*_{\tau2}|\nu_2\rangle + U^*_{\tau3}|\nu_3\rangle.
$$

We can make any matrix element real rephasing a flavor eigenstate. For example the value $U_{e1} = |U_{e1}|e^{i\varphi_{e1}}$ can be made real (and positive) redefining the electron neutrino state $|\nu_e\rangle \to |\nu_e\rangle e^{i\varphi_{e1}}$, and similarly an entire column $(U_{e\mu}, U_{e\tau}, U_{\mu\tau})$ can be made real redefining the three flavor eigenstates and eliminating 3 (or more in general $n$) phases from the matrix. If the neutrino is a Dirac particle, we are also allowed to rephase the mass eigenstates, and therefore we can make real one line $(U_{\alpha1}, U_{\alpha2}, U_{\alpha3})$ in the matrix eliminating other 2 (or more in general $n - 1$) phases. Therefore for Dirac neutrino the rephasing eliminates $(2n - 1)$ phases from the mixing matrix leaving a total of

$$
N_{\text{Dirac}}^{\text{(physical phases)}} = \frac{n(n + 1)}{2} - (2n - 1) = \frac{(n - 1)(n - 2)}{2}.
$$

(55)
For \( n = 2 \) one has \( N(\text{phases}) = 0 \) and the mixing matrix is real, for \( n = 3 \) one phase is present.

If the neutrinos are Majorana particles, there is less freedom to rephase the fields, since one cannot arbitrarily change the phase of the matter fields \( |\nu_j\rangle \). This can be understood observing that the Majorana mass term in the Lagrangian is of the form \( \nu_L \nu_L \) rather than of the form \( \bar{\nu}_R \nu_L + h.c. \) and so the phases of neutrino mass fields cannot be canceled by a rephasing. Therefore only \( n \) phases can be eliminated from the matrix and one remains with additional \((n - 1)\) “Majorana” phases. These Majorana phases are common to an entire column of the mixing matrix, and for this reason they have no effect on neutrino flavor transitions. This can be understood observing that the Hamiltonian that controls the flavor evolution has the form (in the flavor basis):

\[
H = \frac{1}{2E_\nu} \sum_j U_{\alpha j} U_{\beta j}^\dagger m_j
\]

and any phase that is common to an entire column (defined by the index \( j \)) cancels in flavor evolution. These additional phases can however be important in other processes, in particular in double beta decay.

In the case of three flavors, the mixing matrix can then be parametrized with three mixing angles and one CP violating phase (two additional Majorana phases have no influence on the flavor evolution). The most commonly used parameterization has the form:

\[
U = \begin{pmatrix}
1 & 0 & 0 \\
0 & c_{13} & s_{13} e^{-i\delta} \\
0 & -s_{13} e^{i\delta} & c_{13}
\end{pmatrix}
\begin{pmatrix}
c_{12} & s_{12} & 0 \\
-s_{12} & c_{12} & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\theta_{12} & \theta_{13} & \theta_{23}
\end{pmatrix}
\]

where the mixing angles are denoted \( \theta_{12}, \theta_{13}, \theta_{23} \) and have used the notation: \( c_{jk} = \cos \theta_{jk}, s_{jk} = \sin \theta_{jk} \), and the CP violating phase is \( \delta \).

### 4.3. CP and T violating effects

In general the neutrino oscillation probability can violate the \( CP \) or \( T \) symmetry, that is in general:

\[
P(\nu_{\alpha} \rightarrow \nu_{\beta}) \neq P(\bar{\nu}_{\alpha} \rightarrow \bar{\nu}_{\beta}),
\]

\[
P(\nu_{\alpha} \rightarrow \nu_{\beta}) \neq P(\nu_{\beta} \rightarrow \nu_{\alpha}).
\]

On the other hand the \( CPT \) theorem imposes (in vacuum) the condition:

\[
P(\nu_{\alpha} \rightarrow \nu_{\beta}) = P(\bar{\nu}_{\beta} \rightarrow \bar{\nu}_{\alpha}).
\]

The study of the existence of \( CP \) and \( T \) violation effects in neutrino physics is clearly a fascinating and very important topic.

To see how the oscillation probabilities can violate the \( CP \) and \( T \) symmetries, let us consider
again the general (3 flavors) oscillation probability.

\[
P(\nu_\alpha \rightarrow \nu_\beta) = \left| \sum_j U_{\beta j}^* U_{\alpha j} e^{-i m_j^2 \frac{L}{2E}} \right|^2
\]

\[
= \sum_{j=1,3} |U_{\beta j}|^4 |U_{\alpha j}|^4
\]

\[
+ \sum_{j<k} 2 \text{Re}[U_{\beta j}^* U_{\beta k} U_{\alpha j} U_{\alpha k}] \cos \left( \frac{\Delta m_{jk}^2 L}{2E} \right)
\]

\[
+ \sum_{j<k} 2 \text{Im}[U_{\beta j}^* U_{\beta k} U_{\alpha j} U_{\alpha k}] \sin \left( \frac{\Delta m_{jk}^2 L}{2E} \right),
\]

where we have used the notation

\[
\Delta m_{jk}^2 = m_k^2 - m_j^2.
\]

We can see that in general the oscillation probability is composed of a constant term plus 3 “cosine” terms, and 3 “sine” terms. One can immediately see some symmetry properties of the oscillation probability:

- To obtain the oscillation probabilities for anti-neutrinos (replacing neutrinos with anti-neutrinos correspond to a CP transformation) we have simply to replace the matrix \(U\) with its complex conjugate \(U^*\) in the expression of the probability. After the replacement \(U \rightarrow U^*\) in equation (61) the “sin” terms change sign while the “cos” terms remain invariant. We can therefore see that the oscillation probability can violate CP if and only if the mixing matrix has a non vanishing imaginary part, that is if and only of \(\delta\) is different from the values 0 or \(\pi\).

- The oscillation probability for the the time reversed channel (that is the probability for the transition \(\nu_\beta \rightarrow \nu_\alpha\)) can be obtained from equation (61) simply with the interchange of indices \(\alpha \leftrightarrow \beta\). One can immediately see that under this operation the “sin” terms change sign, and the “cos” terms remain invariant. Therefore as in the previous case the oscillation probability can violate T reversal symmetry if and only \(U\) has a non vanishing imaginary part, or equivalently if and only if \(\delta\) is different from the values 0 or \(\pi\).

- A CPT transformation correspond to the replacement of \(\nu\) with \(\nu\) and the interchange of the initial and final flavors, that is formally with the replacements: \(U \rightarrow U^*\), and \(\alpha \leftrightarrow \beta\). It is easy to see that under these transformations expression (61) remains invariant. Therefore the general expression for the oscillation probability respect CPT invariance as it should.

- Inspection of equation (61) also shows that the coefficients of the “sine” terms vanish when \(\alpha = \beta\). Since it is the “sine” terms that are not symmetric under a CP and T operations, it follows that CP and T violating effects do not exist for the “diagonal” transitions (\(\nu_e \rightarrow \nu_e\), \(\nu_\mu \rightarrow \nu_\mu\), \(\nu_\tau \rightarrow \nu_\tau\)). This is as it should be, since it is necessary to satisfy the CPT theorem.

The expression for the oscillation probabilities can be rewritten in a more compact form using the properties of a unitary matrix \(U U^\dagger = U^\dagger U = I\). The unitarity definition can be written more explicilty as:

\[
\sum_j U_{\alpha j} U_{\beta j}^* = \delta_{\alpha \beta},
\]

\[
\sum_\alpha U_{\alpha j} U_{\alpha k}^* = \delta_{jk}.
\]

Using these relations it is possible to obtain relations between the coefficients of the “sine” and “cosine” terms in the expression for the oscillation probability. The most important result is about the quantities
\( J_{jk}^{\alpha\beta} = -\text{Im} \left[ U_{\alpha j} U_{\alpha k}^* U_{\beta j}^* U_{\beta k} \right] \) that are the coefficients of the “sine” terms in the oscillation probability and control the CP and T violation effects. These quantities have obviously the symmetry properties:

\[
J_{jk}^{\alpha\beta} = -J_{kj}^{\alpha\beta} = -J_{jk}^{\beta\alpha}
\]

and

\[
J_{jj}^{\alpha\alpha} = J_{kj}^{\alpha\alpha} = 0.
\] (66)

With some simple algebra using the unitarity conditions one can also obtain the additional relations:

\[
J_{12}^{\beta} = J_{23}^{\beta} = J_{31}^{\beta},
\] (67)

\[
J_{jk}^{e\mu} = J_{jk}^{e\tau} = J_{jk}^{e\nu}.
\] (68)

It follows that (in the case of three neutrinos) there is a single independent coefficient \( J_{jk}^{\alpha\beta} \) the “Jarlskog parameter”:

\[
J = J_{12}^{e\mu} = -\text{Im}[U_{\epsilon 1} U_{\mu 1}^* U_{\epsilon 2}^* U_{\mu 2}]
\] (69)

\[
= c_{13}^2 s_{12} c_{12} s_{23} c_{23} \sin \delta.
\] (70)

All other non vanishing coefficients (those with \( \alpha \neq \beta \) and \( j \neq k \)) are equal or opposite to \( J \). We can therefore finally write the most general form of the probability for the \( \nu_{\alpha} \rightarrow \nu_{\beta} \) transition (for \( \alpha \neq \beta \)) as:

\[
P_{\nu_{\alpha} \rightarrow \nu_{\beta}} = A_{12}^{12} \sin^2 \left( \frac{\Delta m_{12}^2 L}{4 E_{\nu}} \right) + A_{13}^{23} \sin^2 \left( \frac{\Delta m_{23}^2 L}{4 E_{\nu}} \right) + A_{13}^{13} \sin^2 \left( \frac{\Delta m_{13}^2 L}{4 E_{\nu}} \right)
\]

\[
\pm 8 J \sin \left( \frac{\Delta m_{12}^2 L}{4 E_{\nu}} \right) \sin \left( \frac{\Delta m_{23}^2 L}{4 E_{\nu}} \right) \sin \left( \frac{\Delta m_{13}^2 L}{4 E_{\nu}} \right).
\] (71)

In this equations we have used the unitarity conditions to obtain the result that the constant term in (61) vanishes (it is unity for \( \alpha = \beta \)) and have rewritten the CP and T violating part of the probability using the symmetry relations for “sine” coefficients and the trigonometric identity:

\[
\sin a + \sin b - \sin(a + b) = 4 \sin \left( \frac{a}{2} \right) \sin \left( \frac{b}{2} \right) \sin \left( \frac{(a + b)}{2} \right).
\] (72)

We have also used the notation

\[
A_{\alpha j}^{jk} = -4 \text{Re}[U_{\alpha j} U_{\beta j}^* U_{\alpha k}^* U_{\beta k}] .
\] (73)

The CP and T violation effects are described by the last term that changes sign under a CP or T reversal operation. From the expression (70) of \( J \) as a function of the mixing matrix parameters and the form of the oscillation probability, one can see that in order to have observable CP and T violation effects one needs to have three conditions:

[1] The phase \( \delta \) must be non trivial (that is \( \delta \neq 0 \) and \( \delta \neq \pi \))

[2] All mixing angles must be non vanishing: \( \theta_{12} \neq 0, \theta_{23} \neq 0, \theta_{13} \neq 0 \).

[3] All three oscillations must be “active”, that is all three phases \( \Delta m_{jk}^2 L/4E \) must be appreciably different from zero.

---

8The “trick” to use is to note each \( J \) can be seen as the product of two terms of the form \( U_{\alpha j} U_{\beta j}^* \) that can be expressed as a function of each other using the unitarity constraints such as: \( U_{\epsilon 1} U_{\mu 1} = -(U_{\epsilon 2} U_{\mu 2} + U_{\epsilon 3} U_{\mu 3}) \).
4.4. One Mass scale dominance

The expressions for the oscillation probabilities become simpler in the approximation that one squared mass difference is much smaller than the other two. This is realized of the neutrino masses are “hierarchical”, that is:

\[ m_1 \ll m_2 \ll m_3 . \]  

(74)

In this case one has

\[ |\Delta m^2_{12}| \ll |\Delta m^2_{23}| \simeq |\Delta m^2_{13}| . \]  

(75)

In this situation there is a range of \( E_\nu \) and \( L \) where the “short” oscillations are operating but the “long” ones have not developed (and \( |\Delta m^2_{12}| L/4E \ll 1 \)). The oscillation probability can then be well approximated for non-diagonal transitions (\( \alpha \neq \beta \)) as:

\[
P(\nu_\alpha \rightarrow \nu_\beta) = 4|U_{\alpha 3}|^2 |U_{\beta 3}|^2 \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right),
\]

(76)

that is a form similar to the similar to 2-flavor case. The probabilities oscillate with a single frequency related to the squared mass difference \( \Delta m^2_{13} \simeq \Delta m^2_{23} \). The amplitudes of the probability oscillations for the different transitions depends only on the elements of the third column of the lepton mixing matrix \( U \).

Explicitely the oscillation probabilities are:

\[
P(\nu_e \rightarrow \nu_\mu; L) = 4|U_{e 3}|^2 |U_{\mu 3}|^2 \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right)
\]

\[
= s^2_{23} \sin^2 2\theta_{13} \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right),
\]

(77)

\[
P(\nu_e \rightarrow \nu_\tau; L) = 4|U_{e 3}|^2 |U_{\tau 3}|^2 \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right)
\]

\[
= c^2_{23} \sin^2 2\theta_{13} \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right),
\]

(78)

\[
P(\nu_\mu \rightarrow \nu_\tau; L) = 4|U_{\mu 3}|^2 |U_{\tau 3}|^2 \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right)
\]

\[
= c^4_{13} \sin^2 2\theta_{23} \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right).
\]

(79)

The survival probability for diagonal transitions take also the simple form: (for example for \( \nu_e \)):

\[
P(\nu_e \rightarrow \nu_e) = 1 - [1 - |U_{e 3}|^2] |U_{e 3}|^2 \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right)
\]

\[
= 1 - \sin^2 2\theta_{13} \sin^2 \left( \frac{\Delta m^2_{13}}{4E} L \right).
\]

(80)

(81)

It coincides with the \( \nu_e \) survival probability in the 2-flavor case with the substitutions: \( \Delta m^2 \rightarrow \Delta m^2_{13} \) and \( \theta \rightarrow \theta_{13} \).

Another limiting case is relevant for very long baseline reactor experiments (such as KamLand),
and corresponds to the situation
\[ \frac{\Delta m_{13}^2}{2E} L \simeq \frac{\Delta m_{23}^2}{2E} L \gg 1. \] (82)

(that is very fast oscillations due to the mass squared differences \(\Delta m_{13}\) and \(\Delta m_{23}\) with averaged effect). The \(\nu_e\) survival probability is then:
\[ P(\nu_e \rightarrow \nu_e) \simeq c_{13}^4 P + s_{13}^4, \] (83)
with
\[ P = 1 - \sin^2 2\theta_{12} \sin^2 \left( \frac{\Delta m_{21}^2}{4E} L \right), \] (84)

that is the \(\nu_e\) survival probability in the 2-flavor case with the mass squared difference \(\Delta m^2 = \Delta m_{21}^2\) and mixing angle \(\theta = \theta_{12}\).

5. QUANTUM MECHANICS OF NEUTRINO OSCILLATIONS

The phenomenon of flavor oscillations is a fascinating demonstration of quantum mechanical effects in the macroscopic world, and a detailed discussion of its origin can illustrate (and also requires) several important quantum mechanical concepts like coherence, decoherence, wave packets and so on. In this lectures we are concentrating on the phenomenology of the neutrino oscillations and possible interpretations, so we do not have enough space for a careful analysis of these more theoretical aspects. In particular, the derivation of the flavor oscillation formula that we have outlined before, based on the assumption that neutrinos are emitted with a well defined momentum, is simple and elegant, and gives the correct answer for essentially all possible practical applications, however it is not theoretically very solid, it is also not correct in all possible circumstances, and raises several conceptual questions: for example the assumption of a mass independent momentum for neutrinos is in general not correct, and the concept of creation and detection “points” is problematic in a quantum world with an uncertainty principle \(\Delta p \Delta x \geq 1\). A better theoretical treatment of neutrino flavor transitions requires a wave packet treatment of the neutrinos, or more rigorously a quantum field approach. A detailed description of these problems is beyond the scope of these lectures and we refer to [4,37,40] for more discussion and additional references. We include however here a brief discussion to illustrate a few most important points.

5.1. Energy and Momentum of Oscillating Neutrinos

The assumption that the neutrinos are produced with a momentum independent from their mass that was used in our (standard) derivation of the oscillation formula is in general not valid. To have a concrete example, let us consider the neutrinos produced in the decay \(\pi^+ \rightarrow \mu^+ \nu_\mu\) in the rest frame of the parent pion. In this frame, a massless neutrino has an energy (and momentum)
\[ E = p = \frac{m_\pi}{2} \left( 1 - \frac{m_\mu^2}{m_\pi^2} \right). \] (85)

If the neutrino has mass \(m_j\) the energy and momentum are modified, and can be obtained exactly from 4-momentum conservation:
\[ E_j = E + \frac{m_j^2}{2m_\pi} = E + \frac{m_\mu^2}{2E} \left( 1 - \frac{m_\mu^2}{m_\pi^2} \right), \] (86)
\[ p_j = \sqrt{E_j^2 - m_j^2} \simeq E - \frac{m_j^2}{2E} \left( 1 + \frac{m_\mu^2}{m_\pi^2} \right). \] (87)
In these equations we have neglected terms of order \( m^4_j / E^3 \) or higher, and \( E \) is always given by equation (86) Note that both the momentum and the energy of the emitted neutrino depend on its mass. More in general neutrinos produced in different processes will have an energy and a momentum that depend on their mass and can be well approximated by the expressions:

\[
E_j = E + \frac{m_j^2}{2E} \xi, \quad p_j = E - \frac{m_j^2}{2E} (1 - \xi),
\]

where \( E \) is the energy calculated assuming \( m_\nu = 0 \) and \( \xi \) is a dimensionless parameter of order unity that depends on the process we are considering\(^9\), in the example that we have discussed \( \xi = (1 - m_\nu^2 / m_\pi^2) \approx 0.43 \). The expression for flavor transitions will however turn out to be independent from the value of \( \xi \).

The wavefunction of the neutrino state at production can then be represented as the superposition of neutrinos with different masses \( m_j \) each with appropriate momentum \( p_j \) and energy \( E_j \), and amplitudes given by the mixing matrix. A neutrino produced at time \( t_0 \) with flavor \( \nu_\alpha \) has then the wavefunction:

\[
|\psi(x,t_0)\rangle = \sum_j U^*_{\alpha j} |p_j; \nu_j\rangle = \sum_j U^*_{\alpha j} |p_j\rangle |\nu_j\rangle = \sum_j U^*_{\alpha j} e^{ip_j(x-x_0)} |\nu_j\rangle,
\]

where we have represented the the eigenstate of momentum \( |p_j\rangle \) as a plane wave: \( |p_j\rangle = e^{ip_j(x-x_0)} \) with \( x_0 \) the position of \( \pi \) decay. Using the Schrödinger equation for time evolution one can easily write the wavefunction at a later time \( t \) as:

\[
|\psi(x,t)\rangle = \sum_j U^*_{\alpha j} e^{ip_j(x-x_0)} e^{-iE_j(t-t_0)} |\nu_j\rangle.
\]

If we now want to see what is the probability to find a neutrino with flavor \( \beta \) at a distance \( L = x - x_0 \) from the decay point and a time \( T = t - t_0 \) from the decay point we have to compute:

\[
P_{\nu_\alpha \rightarrow \nu_\beta}(L,T) = |\langle \nu_\beta |\psi(x,t)\rangle|^2 = \left| \sum_k U_{\beta k} \langle \nu_k |\psi(x,t)\rangle \right|^2 = \sum_j |U_{\beta j}|^2 |U_{\alpha j}|^2 + \sum_{k>j} 2 \text{Re} \left[ U^*_{\alpha k} U_{\alpha j} U_{\beta k} U^*_{\beta j} e^{i(p_k-p_j)L-i(E_k-E_j)T} \right].
\]

This expression is identical to what we have obtained in the with the assumption that the neutrinos are emitted with a fixed (mass independent) momentum, except that the phase \( \varphi_{jk} \) associated to the product \( U^*_{\alpha k} U_{\alpha j} U_{\beta k} U^*_{\beta j} \) is now:

\[
\varphi_{jk} = (p_k - p_j) L - (E_k - E_j) T \simeq [(p_k - p_j) - (E_k - E_j)] L
\]

\[
= \frac{1}{2E} \left[-(m_k^2 - m_j^2)(1 - \xi) + (m_k^2 - m_j^2)\xi\right] L = \frac{(m_k^2 - m_j^2)}{2E} L.
\]

We can see that the phases (and therefore the expressions for the oscillation probabilities) do not depend from the quantity \( \xi \), and we have exactly the same result obtained before with a derivation based on the fixed momentum assumption.

\(^9\)The expressions (88) respects the condition that the neutrino is on a mass shell \( E_j^2 = p_j^2 + m_j^2 \) up to higher order corrections of order \( m_j^4 / E^3 \).
5.2. Momentum resolution in the \( \nu \) measurement

The phenomenon of flavor oscillation of neutrinos, that is the time (or pathlength) dependence of the probability to find a certain flavor is the consequence not only of mixing, but also of the fact that the neutrino masses are so tiny, that experimentally it is extremely difficult to obtain a resolution that is sufficiently good to determine the neutrino mass, therefore the different amplitudes (corresponding to the different masses) have to be summed coherently, and the quantum interference effects result in the space-time flavor oscillations.

It is instructive to note that in principle the situation could be different, and one can imagine a gedanken experiment measurement capable to determine the mass of an observed neutrino. As an illustration let us consider again the \( \nu_\mu \) produced in the decay of a \( \pi^+ \) at rest. As discussed before the wavefunction of the final state is a superposition of neutrinos of different mass, each having a well defined different momentum. It is in principle possible to perform a measurement with a resolution smaller that the difference in momentum between different mass eigenstates: \( \Delta p < |p_k - p_j| \). In this case however, according to the Heisenberg uncertainty principle the measurement cannot be localized in space better than \( \Delta x \gtrsim 1/\Delta p \). This has the consequence that:

\[
\Delta x \gtrsim \frac{1}{\Delta p} \gtrsim \frac{1}{|p_k - p_j|} \simeq \frac{2 E}{|m_k^2 - m_j^2|} \simeq L_{\text{osc}}^{jk}. \tag{93}
\]

In other words, if the resolution is sufficiently good to determine the mass of the neutrino, then the uncertainty in the position of the observation must be larger than the oscillation length, and no quantum interference can be observed. Therefore, in principle, an extremely precise measurement of the momentum of the neutrino can resolve three distinct values of the momentum \( p_1, p_2 \) and \( p_3 \), each corresponding to the three neutrino mass eigenvalues. The three results are obtained each with a probability \( P_j = |U_{\mu j}|^2 \) that is independent from the neutrino pathlength. The measurement of the neutrino momentum can be performed detecting charged current interactions, and each neutrinos detected with a certain momentum will be in general found not to have a well defined flavor, and the probability to measure the flavor \( \alpha \) for neutrinos with momentum \( p_j \) (and therefore mass \( m_j \)) will be \( P^{(j)}_{\alpha} = |U_{\mu j}|^2 |U_{\alpha j}|^2 \), again independently from the distance between the pion decay and the position of the detector.

This argument explains why in the quark sector one can measure the mixing of weak eigenstates, for example observing that a \( u \) quark makes weak transitions to a state \( d' \) that is a linear combination of the quarks \( d, s \) and \( b \) with measurable coefficients, however one does not observe quark oscillations\(^{10}\). In fact measuring the flavor of a quark is exactly equivalent to the measurement of its mass, and as in the case of neutrinos, resolving the mass destroys the oscillations. Of course in the quark case the mass measurement is enormously easier because of the very large mass differences between different flavors, and viceversa it would be extremely hard to identify experimentally the weak eigenstates \( d', s' \) and \( b' \).

5.3. Wave–packet treatment

The localization of the neutrino source and detection points implies that the neutrinos cannot be described as plane waves, but they must be described as wave packets. The size of the wave packet is determined by the size of the region within which the neutrino production process is localized. Obviously because of the uncertainty principle it follows that even a mass eigenstates cannot be described as having a well defined 3-momentum. The wave–packet treatment is discussed in more detail for example in the textbook of Kim and Pevsner [4]. The introduction of the wave–packet for neutrino also leads to the concept of the “coherence length” for the oscillations.

\(^{10}\)Oscillations are observed between particle/anti-particle pairs such as \( K^0 \leftrightarrow \overline{K}^0 \), \( D^0 \leftrightarrow \overline{D}^0 \) or \( B^0 \leftrightarrow \overline{B}^0 \).
5.4. Coherence Length

It is important to discuss the concept of the coherence length for neutrino oscillations. Neutrinos with different masses will have different velocities given by:

\[ v_j = \frac{p_j}{E_j} \simeq 1 - \frac{m_j^2}{2E_j^2}. \]  

(94)

The neutrino emitted in a certain process can be seen as a wave packet with a size \( \sigma_x \) that is determined by the size of the region where the neutrino production is localized. One can approximate the wave packet as a gaussian of size \( \sigma_x \), and correspondingly the momentum distribution of the neutrino will also be a gaussian with width \( \sigma_p = 1/\sigma_x \) and centered at a value \( p_j \). The wave packet of a neutrino created with a certain flavor, will be in general the superposition of three overlapping wave packets corresponding to the three \( \nu \) masses. The three components of the wave functions will propagate with different group velocities \( v_j' = \frac{1}{m_j^2} (2E_j^2) \), and therefore during propagation the different components will separate. The overlap between two components \( \nu_j \) and \( \nu_k \) will remain significant only if the separation of the centers of their wave packets is smaller that \( \Delta x \sim 2\sigma_x \); this will be correct only for a pathlength smaller than

\[ L < L_{coh}^{jk} = \frac{2\sigma_x}{|v_j - v_k|} \simeq \frac{4\sqrt{2} E^2 \sigma_x}{|\Delta m_{jk}^2|}, \]  

(95)

(the factor \( \sqrt{2} \) at this point is arbitrary). For \( L > L_{coh} \) the wave packets that corresponds to the different neutrinos are separated (see Fig. 6) the coherence between the different neutrinos is lost and oscillations are not present any more. An observer at a distance \( L \gg L_{coh} \) from a neutrino source that generates neutrinos of flavor \( \alpha \) will detect neutrinos of all flavors, but with probabilities independent from \( L \).

The coherence length is in practical not important except for the neutrinos produced in distant supernovae.

![Fig. 6: Scheme of the decoherence of a neutrino wave packet during propagation.](image)

Fig. 6: Scheme of the decoherence of a neutrino wave packet during propagation. The components of the packet that corresponds to the masses \( m_1 \) and \( m_2 \) propagate with different group velocity \( v_1 \) and \( v_2 \) and after a distance of order \( L_{coh} \sim \sigma_x/|v_2 - v_1| \) the packets do not overlap any more and the phenomenon of oscillation stops. An observer will receive separate neutrino pulses that corresponds to the different neutrino masses.

5.5. More General Formula

The effects that we have discussed above can be included in the formula for the probability \( \nu_\alpha \rightarrow \nu_\beta \) using a wave packet, or more rigorously a quantum field theory treatment with the result [40]:

\[
P_{\nu_\alpha \rightarrow \nu_\beta}(x) = \sum_j |U_{\beta j}|^2 |U_{\alpha j}|^2 + \sum_{k > j} 2Re \left[ U_{\alpha k}^* U_{\alpha j} U_{\beta k}^* U_{\beta j} e^{i2\pi \frac{x}{L_{osc}^{\nu \nu}} e^{-\left(\frac{x}{L_{coh}^{\nu \nu}}\right)^2}} \right] e^{-2\pi^2 \xi^2 \left(\frac{x}{L_{osc}^{\nu \nu}}\right)^2}, \]

(96)

where \( L_{osc}^{\nu \nu} \) is the usual oscillation length and \( L_{coh}^{\nu \nu} \) was defined in (95), and \( \sigma_x \) is the combination of the spatial uncertainties in the neutrino production and observation points. This formula
is identical to the “standard” one that we have discussed before except for the last two factors, that are both unity when \( x < L_{\text{coh}} \) or \( \sigma_x \ll L_{\text{osc}} \). The first one is related to the decoherence of the neutrinos that we have just discussed, and the second one is related to the simple condition that the neutrino must be localized better than an oscillation length. In all the problems that we will consider in this work \( L_{\text{coh}} \) is much longer than the neutrino pathlength, and \( \sigma_x \) can be safely assumed to much shorter that the oscillation lengths, and therefore the standard formula that neglects the effects discussed here is valid to a very good approximation.

5.6. Do charged leptons oscillate?

In charged current interactions the neutrinos are always produced in association with a charged lepton, and it is natural to ask the question if some sort of oscillatory behaviour is also associated to these charged leptons. For example since in the decay \( \pi^+ \rightarrow \nu_\mu + \mu^+ \) the probability to observe a \( \nu_\mu \) is a function of the distance \( L \) from the decay point, should the probability of detecting a muon also have a non trivial \( L \) dependence that is the combination of an exponential behaviour (the muon is an unstable particle) with some oscillations? In other words: can “muon oscillations” exist? The answer is no, they cannot exist. The probability of detecting a charged lepton at a certain distance from its production point (in vacuum) is either a constant (for \( e^- \)) or a simple exponential (for \( \mu^- \) and \( \tau^- \)). In the literature there are actually claims to the contrary [41], that however have been solidly refuted [37,40].

It is easy to convince oneself of the non existence of charged lepton oscillations with considerations that are similar to those outlined before. The \( e, \mu \) and \( \tau \) states are eigenstates of mass, and therefore the identification of a charged lepton is exactly equivalent to a measurement of its mass. As in the case of quarks, the coherence in the production is immediately lost, and no oscillations are present. Note also that the charged lepton equivalent of the neutrino weak eigenstates are those linear combinations of \( e, \mu \) and \( \tau \) that are produced in association with a the neutrino mass eigenstates \( \nu_1, \nu_2 \) or \( \nu_3 \), and experimentally is is essentially impossible to measure these states.

It would be instructive to consider the amplitude of the evolution of a two particle final state composed of a neutrino and a charged lepton (as for example in \( \pi^\pm \) decay) and study what correlations can exist between measurements of the two particles. This interesting study is in fact an example of the Einstein, Rosen, Podolsky (EPR) phenomenon, and mathematically it follows analogous steps as for the study of correlations in the processes \( \phi \rightarrow K\bar{K} \text{ or } \Upsilon(4s) \rightarrow B\bar{B} \) that are at the basis of \( CP \) violaton studies in the quark sector. A careful analysis shows that, as expected, no muon oscillations are visible also in the case of measurements of both neutrino and charged lepton (see [40]).

6. NEUTRINO OSCILLATIONS in MATTER

The discussion of the previous Section was relevant for neutrinos propagating in vacuum, however in many (actually in most) circumstances, neutrinos propagate in a medium filled with material. For example solar neutrinos are created in the center of the sun, and have to travel across a lot of solar material, many of the atmospheric neutrinos have to cross a thick layer, or even the entire Earth before reaching the detector, even reactors and accelerator neutrinos, can have to cross thick layers of material. The presence of matter can have a profound effects on the oscillation probabilities [42,43].

6.1. Neutrino Effective Potential in matter

Neutrinos have an extremely feeble interaction with matter, and therefore one could naively expect that the presence of matter should have a negligible effect on their propagation, however this is not so because one must consider the coherent effect of forward scattering from many particles on the propagation of
a neutrino. The effect of coherent forward scattering can be described as an index of refraction for neutrinos in matter, or equivalently as the presence of an Effective Potential due to the presence of matter. Ordinary matter is composed of electrons, protons and neutrons, and the effective potential receives contributions from all target particles. The interaction can proceed with $Z^0$ exchange, resulting in the effective potentials:

\[ V_{\nu e} = V_{\nu \alpha} = \frac{-\sqrt{2}}{2} G_F N_e, \]
\[ V_{\nu p} = V_{\nu p} = \frac{\sqrt{2}}{2} G_F N_p, \]  
\[ V_{\nu n} = V_{\nu n} = \frac{-\sqrt{2}}{2} G_F N_n. \]  
\[ (97) \]

The effective potential for $\nu_e$ is different because the scattering (see Fig. 7) can also proceed with $W$ exchange:

\[ V_{\nu e} = V_{\nu e}^Z + V_{\nu e}^W = \frac{-\sqrt{2}}{2} G_F N_e + \sqrt{2} G_F N_e. \]  
\[ (98) \]

The total effective potential is the sum of the three contributions (for example $V_{\nu e} = V_{\nu e} + V_{\nu p} + V_{\nu n}$). The crucial role is played by the difference of potential between the $\nu_e$ and $\nu_\mu$:

\[ V \equiv V_{\nu e} - V_{\nu e} = V_{\nu e} - V_{\nu e} = +\sqrt{2} G_F N_e. \]  
\[ (99) \]

This potential difference is proportional to the electron density $N_e$, and has the numerical value:

\[ V = +\sqrt{2} G_F N_e = +\sqrt{2} G_F \frac{\rho}{M_p} Y_e = 3.8 \times 10^{-14} \left( \frac{\rho}{1 \text{ g cm}^{-3}} \right) \left( \frac{Y_e}{1} \text{ eV} \right), \]  
\[ (100) \]

(where $Y_e$ is the electron number per nucleon). It is important to note that for anti-neutrinos the effective potential reverse it sign:

\[ V_{\bar{\nu}_\alpha} = -V_{\nu_\alpha}. \]  
\[ (101) \]

Fig. 7: Feynman diagrams for the forward elastic scattering of a $\nu$ from a particle of matter. (a) $W$-exchange-induced scattering from an electron, possible only for $\nu_e$. (b) $Z$-exchange-induced scattering from an $e^-$, $p$ or $n$. possible for all flavors with equal amplitude.
6.2. Flavor Evolution Equation

In the presence of matter the flavor evolution of neutrinos is described by an effective Hamiltonian that is the sum of a free Hamiltonian plus a matter induced term that includes the effective potential and that therefore changes sign for neutrinos and antineutrinos:

\[
\mathcal{H}(\nu) = \mathcal{H}_0 + \mathcal{H}_m, \\
\mathcal{H}(\bar{\nu}) = \mathcal{H}_0^* - \mathcal{H}_m.
\]  

Writing more explicitly for neutrinos one has:

\[
\mathcal{H}(\nu) = U \begin{bmatrix} E_1 & 0 & 0 \\ 0 & E_2 & 0 \\ 0 & 0 & E_3 \end{bmatrix} U^\dagger + \begin{bmatrix} V_e & 0 & 0 \\ 0 & V_\mu & 0 \\ 0 & 0 & V_\tau \end{bmatrix}.
\]

This Hamiltonian enters a Schrödinger equation that controls the flavor evolution:

\[
i\frac{d}{dx} \nu_\alpha = \mathcal{H} \nu_\alpha.
\]

The subtraction from the Hamiltonian of a term proportional to the unit matrix, clearly does not change the flavor evolution, therefore all that counts for the matter term is a matrix of form \( \mathcal{H}_m = \text{diag}[V, 0, 0] \). It should be clear that when \( V \) is comparable to the energy differences \( |E_j - E_k| \approx |\Delta m^2_{jk}|/(2E_\nu) \) the \( \nu \) flavor evolution is strongly modified.

It can be useful to note that the potential can be seen as having the effect of changing the mass of the neutrino. In fact we can write:

\[
E - V = \sqrt{p^2 + m^2} \simeq p + \frac{m^2}{2p}
\]

and approximating \( p \simeq E \) one finds that the effect of the potential is equivalent to a shift in the squared mass:

\[
m^2 \rightarrow m^2 + 2EV.
\]

Quantitatively the potential difference is therefore equivalent to a shift in the squared mass difference:

\[
(\Delta m^2)_{\text{matter}} = \pm 2E_\nu V \simeq \pm 0.76 \times 10^{-4} \left( \frac{\rho}{\text{g cm}^{-3}} \right) \left( \frac{E_\nu}{\text{GeV}} \right) \left( \frac{V_e}{0.5} \right) \text{eV}^2.
\]

Note that electron neutrinos (anti-neutrinos) acquire effectively a larger (smaller) squared mass. Of course a more complete discussion must include mixing (see below).

In general the matter term in equation (105) can vary along the neutrino trajectory. If the density is constant however the flavor evolution can be solved analytically calculating the eigenvalues and eigenvectors of the (time independent) Hamiltonian.

6.3. Two flavor case

In the following we will discuss the simpler case where \( \nu_e \) is mixed with \( \nu_\mu \) and \( \nu_\tau \), and study the structure of the eigenvalues and eigenvectors in matter as a function of the potential \( V \). This will turn out to be of more than academic interest, because it applies to a good approximation to the oscillation of solar neutrinos (where in fact \( \nu_e \) is mixed with a linear combination of \( \nu_\mu \) and \( \nu_\tau \)).
Hamiltonian in vacuum

For pedagogical purposes let us reconsider the free Hamiltonian in the case of two neutrino species. Writing explicitly the mixing matrix (with the compact notation $c = \cos \theta$, $s = \sin \theta$) one has:

$$\mathcal{H}_0 = U \text{ diag}[E_1, E_2] U^\dagger$$

$$= \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

$$= \frac{1}{2E} \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} + \lambda I$$

$$= \frac{\Delta m^2}{4E} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} + \lambda' I$$

(109)

(The terms $\lambda I$ or $\lambda' I$ are proportional to the unit matrix and can be dropped.)

Hamiltonian in matter

We can now add the matter Hamiltonian, that we can recast in the form:

$$\mathcal{H}_\text{mat} = \begin{bmatrix} V_e & 0 \\ 0 & V_\mu \end{bmatrix} = \begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix} + \lambda'' I = \frac{V}{2} \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} + \lambda''' I,$$

(109)

where the terms proportional to the unit matrix $I$ can be dropped because are irrelevant for the flavor evolution. Adding the matter potentials to the free Hamiltonian one obtains:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_\text{mat} = \frac{\Delta m^2}{4E} \begin{bmatrix} -\cos 2\theta + \xi & \sin 2\theta \\ \sin 2\theta & +\cos 2\theta - \xi \end{bmatrix},$$

(110)

where we have defined

$$\xi = \frac{2VE}{\Delta m^2} = \frac{2\sqrt{2} G_F N_e}{\Delta m^2}.$$

(111)

It is now straightforward to diagonalise the matter hamiltonian, comparing with the form of the free one. In fact dividing and multiplying by $\sqrt{\sin^2 2\theta + (\cos 2\theta - \xi)^2}$ we can rewrite:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_\text{mat} = \frac{\Delta m^2}{4E} \begin{bmatrix} -\cos 2\theta + \xi & \sin 2\theta \\ \sin 2\theta & +\cos 2\theta - \xi \end{bmatrix}$$

$$= \frac{\Delta m^2}{4E} \begin{bmatrix} -\cos 2\theta_m & \sin 2\theta_m \\ \sin 2\theta_m & +\cos 2\theta_m \end{bmatrix},$$

(112)
Fig. 8: This figure shows a scheme of the relation between the flavor, mass and propagation eigenstates in the case of two neutrino mixing. For neutrinos (antineutrinos) propagating in matter the propagation eigenstates do not coincide with the mass eigenstates. We have labeled the mass eigenstates so that $m_1 < m_2$ (and therefore $\Delta m^2 = m_2^2 - m_1^2 > 0$). With this convention for the labeling of the mass eigenstates, for neutrinos $\theta^\nu_m > \theta$, while for antineutrinos $\theta^\nu_\bar{m} < \theta$. In the limit of very large matter effects ($(E \times V) \gg \Delta m^2$) $\theta^\nu_m \to \pi/2$ and $\theta^\nu_\bar{m} \to 0$. In the case shown (where $\theta = 25^\circ$) $\theta^\nu_m$ becomes $45^\circ$ for a certain value of $(E \times V)$ and a “MSW resonance” is present for neutrinos (and not anti-neutrinos). For $\theta > 45^\circ$ the resonance exists for anti-neutrinos.

where

$$\sin^2 2\theta_m = \frac{\sin^2 2\theta}{\sin^2 2\theta + (\cos 2\theta - \xi)^2}$$

and

$$(\Delta m^2)_{\text{eff}} = \Delta m^2 \times \sqrt{\sin^2 2\theta + (\cos 2\theta - \xi)^2}. \quad (114)$$

In matter the neutrino and anti-neutrino mixing are different because $\nu$ and $\bar{\nu}$ have opposite effective potentials. The eigenvalues/eigenvectors solutions for anti-neutrinos can be obtained with the replacement $V \to -V$, that is $\xi \to -\xi$. The situation is illustrated in Fig. 8.

The MSW resonance

It is very instructive to study the behaviour of effective mixing parameters in matter as a function of the quantity $\xi = 2VE/\Delta m^2$ (see Fig. 9). Note that $\xi$ is proportional to the density $V$ and to the neutrino energy $E$. Without loss of generality it is possible to assume that $\Delta m^2$ is positive, if the mixing angle $\theta$ can vary in the interval $\theta \in [0, \pi/2]$. In this case $\theta = 0$ corresponds to the situation where the lightest neutrino state is a pure $\nu_e$ and $\theta = \pi/2$ to the situation where the lightest state is a $\nu_{\mu,\tau}$\textsuperscript{11}. With this convention $\xi$ is always positive for neutrinos always negative for antineutrinos.

\textsuperscript{11}For two flavor mixing it is also possible to use a different convention, reducing the range of variability of $\theta$ in the range $[0, \pi/4]$, and considering both positive and negative $\Delta m^2$. In this convention the $\nu_1$ is always the state with the largest overlap with $\nu_e$; while in the other convention $\nu_1$ is the eigenstate with the lowest mass.
The effective mixing parameter in matter, as shown in Fig. 9, can be expressed as a function of 
\[ \sin^2 2\theta_{\text{eff}} = (2 E V) / \Delta m^2 \]
where \( \Delta m^2 \) and \( V \) are energy differences and potential differences, respectively.

The main features of the effective mixing in matter in Fig. 9, can be readily understood from equation (113):

1. The low density limit can be obtained for \( \xi \to 0 \), it is easy to see that one recovers the vacuum case.
2. The limits \( \xi \to \pm \infty \) correspond to very high energy (or very large density) for neutrinos or antineutrinos. The effective mixing parameters become:

\[
\sin^2 2\theta_{\text{eff}} = \sin^2 2\theta \left( \frac{\Delta m^2}{2E} \right)^2 \to 0, \\
(\Delta m^2)_{\text{eff}} \to \Delta m^2 \times \xi = \pm 2E \xi.
\]

3. The mixing parameter \( \sin^2 2\theta_{m} \) becomes unity (that is \( \theta_{m} \) becomes \( \pi/4 \)) for the condition:

\[
\xi = \cos 2\theta, \\
2E \cos \theta = \Delta m^2 \cos 2\theta.
\]

4. The width of the “resonance region”, that is the range of \( \xi \) for which \( \sin^2 2\theta_{m} \) is larger than 0.5 is \( \xi \in [\cos 2\theta - \sin^2 \theta, \cos 2\theta + \sin^2 \theta] \) and shrinks with decreasing \( \theta \).

The relation between flavor, mass and propagation eigenstates can be pictured geometrically observing that with increasing density the mixing angle \( \theta_{m} \) for neutrinos “rotates” monotonously anti-clockwise from the vacuum value \( \theta \) (for a vanishingly small density) toward increasing values, reaching for very large densities the asymptotic value \( \theta_{m} \to \frac{\pi}{2} \). For antineutrinos the mixing angle \( \theta_{m} \) rotates in the opposite direction toward the asymptotic value \( \theta_{m} \to 0 \). Clearly the mixing angle in matter must become “maximal” (that is take the value \( \theta_{m} = 45^\circ \)) for neutrinos or antineutrinos, for a particular value of the product \( E \nu N_e \). This condition will be present for neutrinos when \( \theta < 45^\circ \) that is when the lightest mass state is predominantly a \( \nu_e \) as it is naively expected “normal” ordering of the neutrino masses, or for anti-neutrinos if \( \theta > 45^\circ \) that is for an inverted hierarchy of the \( \nu \) masses.

[12] In the other convention of having the angle \( \theta \) limited to the region \( \theta \in [0, \pi/4] \), one would conclude that the MSW resonance is present for neutrinos for \( \Delta m^2 > 0 \) or for anti-neutrinos for \( \Delta m^2 \) negative. The statements have of course the identical physical meaning.
Figure 9 shows a plot of the effective mixing angle $\theta_m$ and of the effective mixing parameter $\sin^2 2\theta_m$ for neutrinos and anti-neutrinos. One can see that the mixing parameter has a “resonance like” behaviour around the value $\xi = \cos 2\theta$ where the mixing becomes maximal. The “MSW” resonance (from the initials of Mikheyev, Smirnov and Wolfenstein [42] that discovered it) has a width (half width at half maximum) $\delta \xi = \sin 2\theta$ that becomes narrower with decreasing $\sin 2\theta$.

In Fig. 9 we plot an example of the effective squared mass eigenvalues as a function of the product $2VE$. This illustrates that at the resonance $\Delta m^2_{\text{eff}}$ has a minimum and and becomes $\Delta m^2_{\text{eff}} = \Delta m^2 \sin 2\theta$. The $\Delta m^2_{\text{eff}}$ at the resonance vanishes for $\theta = 0$ or $\theta = 90^\circ$. This can be easily understood observing that in the case of no mixing ($\theta = 0^\circ$) the effective squared masses in matter are simply $M_{1,2}^2 = m_{\nu_e}^2 + 2EV$ and $M_{2,1}^2 = m_{\nu_\mu}^2$. In this case the “resonance” corresponds simply to a the level crossing at $2EV = (m_{\nu_\mu}^2 - m_{\nu_e}^2)$. Below the “resonance” or crossing point the heaviest propagation eigenvector is a $\nu_\mu$, above the crossing point it becomes a $\nu_e$.

In the presence of a small non vanishing $\theta$, there is no more a “crossing” of the levels, however something very similar happens in the flavor content of the propagation eigenstates. For example the propagation eigenstates that correspond to the largest (heaviest) eigenvalue correspond to a nearly pure $\nu_\mu$ for for small $\xi$ and to a nearly pure $\nu_e$ for large $\xi$. This phenomenon is possibly the key mechanism in the oscillation of solar neutrinos.

### 6.4. Variable density

In many cases of interest neutrinos propagate in non homogeneous matter. Examples of this problem are:

1. Solar neutrinos that are created near the center of the sun, where the density is of order $130 \text{ g cm}^{-3}$, and travel to the surface with the density decreasing approximately exponentially along the trajectory.
2. Atmospheric neutrinos can cross the Earth, where the density is approximately $2.7 \text{ g cm}^{-3}$ near the surface but higher than $12 \text{ g cm}^{-3}$ near the center.
3. Supernova neutrinos have also to cross a large column density of material with a rapidly varying electron density.

In general the transition probabilities can be easily calculated numerically solving the flavor evolution equation:

$$i \frac{d}{dx} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} \mathcal{H}_0 + \mathcal{H}_{\text{mat}}(x) \end{pmatrix} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}$$

$$= \begin{cases} \frac{1}{2E} U \begin{bmatrix} m_1^2 & 0 & 0 \\ 0 & m_2^2 & 0 \\ 0 & 0 & m_3^2 \end{bmatrix} U^\dagger + \begin{bmatrix} V(x) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{cases} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}, \quad (119)$$

where $V(x) = V_e(x) - V_\mu(x) = \sqrt{2} G_F N_e(x)$ is the effective matter potential. The initial condition of the equation can be taken as a flavor eigenstate (for example $|\nu_e\rangle$: $|\nu_{\text{ini}}\rangle = \delta_{\alpha,e}$, and the probability $P_{\nu_e \rightarrow \nu_\alpha}$ is obtained squaring the final neutrino $\alpha$ component.
7. ATMOSPHERIC NEUTRINOS

Atmospheric neutrinos are produced in the collisions of cosmic rays with the Earth’s upper atmosphere. Cosmic rays are relativistic charged particles that reach the Earth with an isotropic flux of approximately 1 particle/(cm$^2$ sec sr). They are produced in astrophysical sources whose nature, after nearly a century of research, is still in question (the most likely sources are the blast waves produced by supernovae explosions), most cosmic rays are protons with contributions of completely ionized nuclei, and smaller components of electrons, anti-protons and positrons. Cosmic rays have a spectrum that to a good approximation follows a power law behaviour: $\phi_A(E) \propto E^{-\alpha}$ with $\alpha \approx 2.7$, extending up to extremely high energies $E \sim 10^{20}$ eV. The galactic magnetic fields trap the cosmic rays for a long (energy dependent) time of order of millions of years, and during this time the directions of the particles are completely scrambled, so that the flux is to a very good approximation isotropic and uniform in time, because time variations in the source intensity are averaged over a long time. Primary cosmic ray interact in the upper atmosphere, with the air nuclei producing secondary particles:

$$p + A_{Air} \rightarrow p, n, \pi^\pm, \pi^0, K^\pm, \ldots$$

that initiate showers in the atmosphere. Neutrinos are abundantly produced in these showers, the dominant source is the decay of $\pi^\pm$, and the subsequent decay of $\mu^\pm$, such as:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\downarrow$$

$$e^+ + \nu_e + \overline{\nu}_\mu$$

(and charge conjugate channels), similarly neutrinos are also produced in the decay of kaons.

For a review on atmospheric neutrinos see [44].

7.1. Robust properties of the predicted fluxes

The detailed calculation of the atmospheric neutrino fluxes is a non trivial problem (for reviews see [45,46]), that requires the following elements:

1. A description of the primary cosmic rays $\phi_A^c(E)$, with fits to the measurements obtained from balloon and satellites experiments.
2. A model of hadronic interactions to compute the multiplicity and energy and angular distributions of the final state particles produced in hadron-nucleus and nucleus-nucleus interactions.
3. A calculation scheme (usually a montecarlo method) to compute the development of the showers taking into account all relevant processes, like the energy losses of charged particles, the competition between interaction and decay for unstable hadrons, and the weak decays of mesons and muons.

Independently from the details of this calculation, and using the assumption that the flavor of the neutrinos does not change during propagation from the creation point to the detector, one can obtain two fundamental and very robust properties of the fluxes:

- The flux of $\nu_\mu + \overline{\nu}_\mu$ is approximately twice as large as the flux of $\nu_e + \overline{\nu}_e$

$$[\phi_{\nu_\mu}(E_\nu, \Omega_\nu) + \phi_{\overline{\nu}_\mu}(E_\nu, \Omega_\nu)] \simeq 2 \times [\phi_{\nu_e}(E_\nu, \Omega_\nu) + \phi_{\overline{\nu}_e}(E_\nu, \Omega_\nu)].$$

- The fluxes of all neutrino species are up-down symmetric:

$$\phi_{\nu_\alpha}(E_\nu, \theta_\nu) = \phi_{\nu_\alpha}(E_\nu, \pi - \theta_\nu).$$
The condition (121) is a simple consequence of equation (120), that tell us that, after the completion of the chain decay, for each $\pi^+$ ($\pi^-$) there is one $\nu_\mu$, one $\nu_\tau$ and one $\nu_e$ ($\overline{\nu}_e$) that have approximately the same average energy. The last remark about the average energy of the neutrinos is necessary to insure that the relation (121) between the neutrino fluxes is valid not only for the total number of neutrinos, after integration over all energies, but also for the differential fluxes at each neutrino energy. One could naively think that the “second generation” neutrinos produced in muon decay will have on average a lower energy that the produced directly in the $\pi^+$ decay, however this is not the case, because in the first decay the muon carries away a large fraction of the pion energy since $m_\mu$ is close to $m_\pi$: $\langle E_{\mu^+} \rangle / E_{\pi^+} = (1 + m_\mu^2 / m_\pi^2) / 2 \simeq 0.787$, and the average energy of the three neutrinos are very similar. Ignoring the energy loss of the muon before decay one has $\langle E_{\nu_\mu} \rangle / E_{\pi^+} = 0.213$, $\langle E_{\nu_\tau} \rangle / E_{\pi^+} = 0.265$, $\langle E_{\overline{\nu}_e} \rangle / E_{\pi^+} = 0.257$. Equation (121) ceases to be valid at high energy ($E_\nu \gtrsim 3$ GeV for vertical neutrinos) when, because of relativistic effects, the muon decay length becomes longer than the thickness of the atmosphere, and muons reach the ground dissipating their energy in ionization.

The prediction of the up-down symmetry (equation (122)) is even more robust, and in fact can be demonstrated as a simple geometry theorem from the (quasi-exact) spherical symmetry of the problem that can be restated as:

[1] The Earth can be well described as a perfect sphere.

[2] The primary cosmic rays are isotropic, that is the fluxes that arrive at different points on the Earth are all equal.

The demonstration is simple. The neutrinos that enter a sphere with zenith angle $\theta_{z}^{\text{down}}$ will exit the sphere with zenith angle $\theta_{z}^{\text{up}} = \pi - \theta_{z}^{\text{down}}$ (see Fig. 10). The absorption of neutrinos inside the Earth is a negligible effect (of order $\sim 10^{-6}$ for vertical $\nu_{\mu,e}$ at 1 GeV), and therefore the number of neutrinos that enter the Earth with a given energy and zenith angle $\theta_{z}$ is equal to the number of neutrinos that exits with zenith angle $\pi - \theta_{z}$ and the same energy:

$$N_{\nu_\alpha}^{\text{in}}(E, \theta_{z}) = N_{\nu_\alpha}^{\text{out}}(E, \pi - \theta_{z}) .$$

(123)

For spherical symmetry (isotropic primary cosmic rays) the neutrino flux is equal for all points on the Earth, and can be related to the total number of crossing neutrinos as

$$\phi_{\nu_\alpha}(E, \theta_{z}) = \frac{N_{\nu_\alpha}^{\text{in(out)}}(E, \theta_{z})}{4\pi R_{\oplus}^2 |\cos \theta_{z}|} ,$$

(124)

where $4\pi R_{\oplus}^2$ is the Earth surface and the factor $|\cos \theta_{z}|$ takes into account the orientation of the surface with respect to the neutrino direction. The equality of the fluxes follows immediately. It is clear that this prediction of up-down symmetry for the neutrino fluxes, offers a gold-plated method to study flavor oscillations. The pathlengths for up-going and down-going neutrinos are very different (by a factor of order $10^3$ for the vertical directions) and in the presence of oscillations one expects that the two fluxes are modified in different ways. Therefore, if oscillations exist, the up-down symmetry will be broken and the effect can be easily observed measuring a difference in the rates of up-going and down-going events.

Since the up-down symmetry (or lack of) plays a crucial role in the evidence for oscillations in atmospheric neutrinos it is worthwhile to consider what is the maximum size of the asymmetry that can be expected in the absence of oscillations. The sources of asymmetry are the following:

[1] The largest violations of spherical symmetry is due to the presence of the geomagnetic field. The field bends the trajectories of the primay particles, so that low rigidity particles cannot reach the surface of the Earth. The effect is stronger at the magnetic equator (where the cosmic ray flux...
Fig. 10: The geometrical origin of the up-down symmetry of the atmospheric neutrino fluxes. The symmetry is present in the absence of neutrino oscillations.

is minimum and vanishes at the magnetic pole (where the cosmic ray is maximum). This effect influences also the lowest energy neutrinos, because high rigidity cosmic rays are not influenced by the geomagnetic field. For a detector placed near the magnetic equator (like Super-Kamiokande in Japan), the effect is a suppression of the down-going hemisphere (where the equatorial field is operating) relative to the up-going one (since the average effect of the geomagnetic field is weaker). The effect is reversed for detectors placed near a magnetic pole.

[2] The Earth is not a perfect sphere, and its general shape can be approximated as an ellipsoid of rotation, however it can be shown that this effect is entirely negligible. More important are those small ‘irregularities’ represented by mountains. The number of neutrinos produced in showers that develop above a mountain is (slightly) reduced because of the reduced thickness of air. Since most (~ 0.7) of the Earth surface is at sea level, while most neutrinos detectors are placed below a mountain (that is used as an absorber), this results in a small reduction of the down-going flux.

[3] A third (smaller effect) is due to the average density profile of the atmosphere that changes with latitude. Since the profile has also seasonal variations, the size of this effect can be constrained by the (non)-observations of seasonal effects.

7.2. Detectors

The event rate for atmospheric neutrinos is of order ~ 0.5 events/(Kton day), and to have appreciable rates one needs very massive detectors. Moreover, to reduce the background due to secondary cosmic rays that at the surface (at sea level) have a flux ~ 200 particles/(m² sec), the detectors have to be placed deep underground. The component of $\gamma$, $e^\pm$ and hadrons can be absorbed with few meters of rock, but to reduce significantly the flux of the penetrating muons, one needs a coverage of order of 1 Km of rock. Detectors for atmospheric neutrinos have been constructed using essentially two techniques: iron calorimeters with a segmentation of order 1 cm (such as Frejus, Nusex and Soudan-2) and Water Cherenkov detectors like Kamiokande, IMB and Super-Kamiokande. Relativistic charged particles moving in water with a velocity larger than $\beta_c = 1/n \approx 0.75$ (where $n \approx 1.33$ is the refractive index of the material) emit Cherenkov radiation; when $\beta \to 1$ the light is emitted in a cone of semi-angle $\theta_C = \cos^{-1} n \approx 42^\circ$. Since water is transparent to the Cherenkov radiation, the photons can be detected with sensitive elements (photo multiplier tubes or PMT’s) placed at the surface of the detector. This method allows the construction of large detectors at a relatively low cost.
The largest sample of data on atmospheric neutrinos has been obtained by the Super-Kamiokande detector (see Fig. 11). The detector is a stainless steel cylinder with 42 meters of height and 39 meters of diameter that contain 50 Ktons of water. A two meters layer of water is used as a veto counter, while the inner volume (32 Kton) can be used to detect neutrino interactions. The surface of the inner volume is instrumented with 11146 PMT’s each one with a diameter of 50 cm that look “inside”, while 1885 PMT’s with 20 cm diameter look “outside” instrumenting the veto detector. The Super-Kamiokande detector has taken data from april 1996 until november 2001, when the implosion of one phototube, started a chain reaction, that destroyed approximately two thirds of the other PMT’s. The detector is currently being repaired and will start taking data again in 2003.

![Fig. 11: The Super-Kamiokande detector.](image)

The most important class of atmospheric neutrino events are quasi-elastic interactions of type:

\[ \nu_\ell + n \rightarrow \ell^- + p , \]
\[ \bar{\nu}_\ell + p \rightarrow \ell^+ + n . \]

(125)

With increasing \( E_\nu \) events with the production of additional particles in the final state also become possible (see Fig. 12).

![Fig. 12: Atmospheric neutrino events detected by the Soudan–2 detector.](image)

The events at the left can be interpreted as \( \nu_\mu + n \rightarrow \mu^- + p \) because of the \( \mu^- \)-like track and the short proton recoil. The next event is the other view of an event of the same type. The next event can be interpreted as a “single shower” event of type \( \nu_e + n \rightarrow e^- + p \) or \( \bar{\nu}_e + p \rightarrow e^+ + n \). The last event is a multiprong \( \nu_\mu \) (or \( \bar{\nu}_\mu \) interaction).
Muons in water lose energy continuously and emit Cherenkov light until they reach the critical velocity, near the end of their range: $L_{\mu} \simeq 5 \text{ EGeV}$ meters. This results in a ‘ring’ of Cherenkov light at the surface of the detector. From the shape and brightness of the ring it is possible to determine the point of creation of the particle and its direction and energy. Electrons in water undergo bremsstrahlung and originate an electromagnetic shower, so that their energy is dissipated in the ionization of several $e^\pm$ particles, each one generating a Cherenkov ring that overlap in a single visible ring that is more ‘fuzzy’ than those produced by muons, because of multiple scattering effects. The difference between the ‘sharp’ ($\mu$-like) and ‘fuzzy’ ($e$-like) rings allows to determine the flavor of the charged lepton (see Fig. 13). The flavor identification is easier when a single ring is present in the event, and for this reason “single ring” events have been selected by SK at some cost in efficiency at higher neutrino energy when the production of additional particles in the final state becomes more important. The results of Super-Kamiokande are shown in Fig. 14. The 6 panels show the results for different classes of events. The first 4 classes of events are “fully contained events”, that is events where the neutrino interacts inside a detector fiducial volume (2 meters away from the detector walls for a fiducial mass of 22.5 Ktons) and no particle exits from the detector. The events are divided into 4 categories according to the flavor identification ($e$ or $\mu$) and the amount of visible energy (smaller or larger than 1.33 GeV). The other two classes of events are throughgoing and stopping upward going muons. These are events generated by the charged current interactions of $\nu_{\mu}$ and $\bar{\nu}_{\mu}$ in in the rock below the detector.

The results of SK Fig. 14 are shown together with a standard model prediction (in the absence of $\nu$ oscillations), and a with a prediction that assumes the existence of $\nu_{\mu} \leftrightarrow \nu_{\tau}$ oscillations. In this case the prediction depends on the two parameter $\Delta m^2$ and $\sin^2 2\theta$, and the lines shown correspond to the best fit values $\Delta m^2 = 2.5 \times 10^{-3} \text{ eV}^2$ and $\sin^2 2\theta = 1$.

Several points appear clearly from an inspection of Fig. 14. The first one is that for $e$-like events, the data and the standard model (no oscillation) prediction are in good agreement with each other, note in particular that the data exhibits the predicted up-down symmetry. On the other hand for all 4 categories of $\mu$-like events there is a significant deficit with respect to the prediction. Moreover these deficits exhibit some very interesting dependence on the zenith angle. The most spectacular effect, and the key for the interpretation of the result is the zenith angle distribution of the multi-GeV $\mu$-like events, where the number of up-going events (with zenith angle $\cos \theta_z < 0$) is only 1/2 of the number of down-going ones ($\cos \theta_z > 0$).
7.3. The $\nu_\mu \leftrightarrow \nu_\tau$ interpretation.

It is simple to understand qualitatively how the assumption of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations can provide a good description of the data. Under this assumption the neutrino fluxes at the detector are related to the no-oscillation ones by:

$$
\phi_{\nu_e} \rightarrow \phi_{\nu_\mu},
$$

$$
\phi_{\nu_\mu} \rightarrow [1 - \langle P_{\nu_\mu \rightarrow \nu_\tau} \rangle] \phi_{\nu_\mu},
$$

$$
\phi_{\nu_\tau} \simeq 0 \rightarrow \langle P_{\nu_\mu \rightarrow \nu_\tau} \rangle \phi_{\nu_\mu},
$$

and similarly for anti-neutrinos. In this equation, the brackets $\langle \rangle$ correspond to an average over the position of the creation points of the neutrinos (and therefore the pathlength). After this average the the oscillation probability will depend on the neutrino energy and direction, with a form that is determined by the two parameters $\Delta m^2$ and $\sin^2 2\theta$:

$$
\langle P_{\nu_\mu \rightarrow \nu_\tau} \rangle = \langle P_{\nu_\mu \rightarrow \nu_\tau} \rangle (E_\nu, \cos \theta; \Delta m^2, \sin^2 2\theta). \quad (126)
$$

Matter effects are ineffective in the case of two-flavor $\nu_\mu \leftrightarrow \nu_\tau$ oscillations and one has the simple formula:

$$
P_{\nu_\mu \rightarrow \nu_\tau} (L, E_\nu) = \sin^2 2\theta \sin^2 \left[ \frac{\Delta m^2 L}{4E_\nu} \right]. \quad (127)
$$

This probability takes some simple form in the limits of large or small $L/E_{\nu_{\mu \tau}}$ (where large or small is defined comparing with $1/\Delta m^2$):

$$
P_{\nu_\mu \rightarrow \nu_\tau} = \begin{cases} 
0 & \text{for } L \text{ small,} \\
1 - \frac{\sin^2 2\theta}{2} & \text{for } L \text{ large.} 
\end{cases} \quad (128)
$$
In the case of large $L$ we have averaged over fast oscillations.

Evidence for the presence of oscillations in the $\nu_\mu \leftrightarrow \nu_\tau$ case can be obtained either observing the appearance of events with a $\tau^\pm$ in the detector, or observing a suppression of the rate of $\mu$-like events. The threshold of charged current interactions for $\nu_\tau$ and $\bar{\nu}_\tau$ is $E^{\text{thresh}}_{\nu_\tau} \approx m_\tau + m^2/2m_\mu \approx 3.5$ GeV, and the cross section is severely suppressed close to the kinematical threshold, therefore the flux of $\nu_\tau$ and $\bar{\nu}_\tau$ generated by the oscillations is very difficult to observe, while the disappearance effects can be more easily detected. Clearly a detectable effect is a suppression of the rate of $\mu$-like events with respect to the no-oscillation prediction or, to eliminate systematic uncertainties, with respect to the $e$-like rate, moreover it can be expected that disappearance effects can be a function of the zenith angle and energy of the neutrinos, reflecting the functional form of the probability (127).

The neutrino pathlength $L$ is very strongly correlated with the zenith angle $\theta_z$ (see Fig. 15). For a first order analysis it is sufficient to observe that neutrinos are produced at a typical height $h \sim 20$ Km, with only a weak dependence on the energy, flavor and zenith angle (see [47] for more discussion) and therefore to a reasonable approximation

$$L \approx -R_\oplus \cos \theta_z + \sqrt{(R_\oplus \cos \theta_z)^2 + 2R_\oplus h + h^2}.$$  \hspace{1cm} (129)$$

This expression ranges from $L \sim h \sim 20$ Km for vertically down-going neutrinos ($\cos \theta_z = 1$) to $L \approx 2R_\oplus + h \sim 12740$ Km for up-going $\nu$ ($\cos \theta_z = -1$). For horizontal neutrinos the pathlength is $L_{\text{hor}} \approx \sqrt{2R_\oplus h} \sim 500$ Km. A natural expectation is that up-going neutrinos, that travel for a longer distance will have larger oscillation probabilities than down-going ones, and therefore the rate of up-going events will be more suppressed.

In order to understand the parameters $\Delta m^2$ and $\sin^2 2\theta$ that can fit the atmospheric neutrino data it is best to look first at the multi-GeV data. Defining “Up” as the up-going rate (with $\cos \theta_{z,\mu} < -0.4$)
and “Down” as the down-going rate (with $\cos \theta_{\pm, \mu} > 0.4$), the ratio is approximately one half:

$$\left( \frac{\text{Up}}{\text{Down}} \right)_{\text{multiGeV}} = 0.54 \pm 0.04 \pm 0.01.$$  \hspace{1cm} (130)

Experimentally one measures the direction of the charged leptons ($e$ or $\mu$) and not of the neutrino, however the two directions are correlated, and the average angle $\langle \theta_{\ell \nu} \rangle$ shrinks with increasing energy $\propto E^{-1}$. For sub-GeV events $\langle \theta_{\ell \nu} \rangle \sim 60^\circ$, the correlation is rather poor, while for multi-GeV events $\langle \theta_{\ell \nu} \rangle \sim 10^\circ$, the correlation is much more strict, and to first approximation one can assume $\theta_{\pm, \mu} \simeq \theta_{\pm, \nu}$. It is then quite easy to see how data and prediction can be reconciled, assuming the presence of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations and choosing appropriately the parameters $\sin^2 2\theta$ and $\Delta m^2$ so that down-going events have negligible suppression while down-going events have the average suppression as in equation (128). A rough estimate of the oscillation parameters is easy to obtain. The up/down ratio allows to estimate the mixing parameter from:

$$1 - \frac{\sin^2 2\theta}{2} \simeq \left( \frac{\text{Up}}{\text{Down}} \right)_{\text{multiGeV}} \sim 0.5,$$  \hspace{1cm} (131)

giving $\sin^2 2\theta \sim 1$. To estimate $\Delta m^2$ one can use the fact that the critical length that separates “long” (no-oscillation) from “short” (averaged oscillations) pathlengths correspond to the horizontal directions:

$$L^* \simeq \frac{\lambda_{\text{osc}}^*}{2} \simeq \frac{2\pi \langle E_\nu \rangle}{|\Delta m^2|} \simeq 1000 \text{ km}.$$  \hspace{1cm} (132)

For multi-GeV events, in first approximation one has $\langle E_\nu \rangle \sim 3$ GeV, with an estimate:

$$|\Delta m^2| \simeq \frac{2\pi \langle hc \rangle}{L^*} \simeq 3 \times 10^{-3} \text{ eV}^2.$$  \hspace{1cm} (133)

A measurement of the average suppression of the $\nu_\mu$ and $\nu_\mu$ flux can be obtained from the ratio $\langle \mu \rangle_{\text{Data}} / \langle \mu \rangle_{\text{MC}}$ between the measured and predicted event rate. In order to reduce the systematic uncertainty in the prediction it is useful to consider the “Double ratio”

$$R = \left( \frac{\langle \mu \rangle_{\text{Data}}}{\langle e \rangle_{\text{Data}}} \right) / \left( \frac{\langle \mu \rangle_{\text{MC}}}{\langle e \rangle_{\text{MC}}} \right).$$  \hspace{1cm} (134)

Since as discussed before see equation (121)) the ratio between $\mu$-like and $e$-like events can be reliably predicted in the double ratio systematic uncertainties such as the absolute normalization of the flux cancel. For multi-GeV events the double ratio measured by SK is:

$$R_{\text{multiGeV}} = 0.68 \pm 0.03 \text{ (stat)} \pm 0.08 \text{ (sys),}$$  \hspace{1cm} (135)

this corresponds to best estimate for the oscillation probability averaged over all zenith angles and energies for this class of events.

The sub-GeV events are nicely consistent with this interpretation. The average oscillation probability can be obtained from the double ratio $R$:

$$R_{\text{subGeV}} = 0.64 \pm 0.02 \text{ (stat)} \pm 0.05 \text{ (sys)}$$  \hspace{1cm} (136)

is close to the one measured for the multi-GeV events, as it is expected. In this case the correlation between the muon and the neutrino direction is less good, and therefore the up/down asymmetry for muon events is less marked, but it still clearly visible.

The results of a detailed fit to the SK data results in the allowed region (at 90% C.L.) shown in
Fig. 16, give the result:

\[
\begin{align*}
|\Delta m^2| & \sim (1.6 - 4.5) \times 10^{-3} \text{ eV}^2, \\
\sin^2 2\theta & \geq 0.90,
\end{align*}
\]

(137)

that corresponds to the qualitative estimate that we have discussed.

![Allowed regions at the 68,90 and 99% confidence level for the $\nu_\mu \leftrightarrow \nu_\tau$ oscillation hypothesis obtained by Super-Kamiokande.]

7.4. Upward-going muons

A different method to detect fluxes of $\nu_\mu$ and $\overline{\nu}_\mu$ is the observation of neutrino induced upward-going muons. In these events the neutrino interaction happens in the rock below the detector, and the $\mu^\pm$ produced in charged current interactions propagate to the detector. In through-going events the muons cross the detector, while in stopping events the muons range out inside the detector volume. Because of the up-down symmetry of the atmospheric neutrino fluxes one expects (in the absence of oscillations, and if the detector is placed at a sufficient depth) an equal number of $\nu$-induced up-going and down-going muon events, however down-going events cannot be separated from the flux of atmospheric muon events, where the muon is produced directly in the atmosphere, and therefore only up-going events can be used to study the neutrino fluxes, with the exception of exceptionally deep detectors. For a threshold energy of order 1 GeV, the flux of $\nu$-induced upgoing events is of order:

\[
\langle \phi_{\mu\nu} \rangle \approx 2 \times 10^{-13} \text{ (cm}^2 \text{ sec sr)}^{-1} \approx 60 \left[(1000 \text{ m}^2) \text{ year sr}\right]^{-1}.
\]

(138)

For a rock overburden of order $h \sim 1$ Km, the atmospheric muon flux is still approximately $10^5$ times larger than the $\nu$-induced one. The two fluxes become comparable for a thickness of approximately 5 Km of rock.

The range of a muon in rock is $R_\mu \approx 1.7 \text{ EGeV}$ meters, therefore high energy neutrinos can produce up-going muons even if they interact far from the detector. For $\nu_\mu$ and $\overline{\nu}_\mu$ with energy larger than $\gtrsim 10$ GeV (the precise number depends of course on the dimension and geometry of the detector) the event rate in up-going muons is larger than the rate of interactions inside the detector. The neutrino energy that contribute to the fluxes of up-going muons is shown in Fig. 17: the median neutrino energy is of order $E_\nu \sim 10$ GeV for stopping events, and $E_\nu \sim 100$ GeV for through-going events. In the presence of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations with $\Delta m^2$ of the order suggested by the contained events one can expect that the
Fig. 17: Event rates as a function of neutrino energy for fully contained events, stopping muons, and through-going muons at SuperKamiokande.

flux of the (lower energy) passing events is suppressed by the averaged factor $1 - \sin^2 2\theta$ for all zenith angles except for directions very close to the horizontal plane, while for the higher energy through-going events the suppression should be smaller, and with a non trivial dependence on the zenith angle, being largest for the vertical direction.

Up-going muons have been the first type of atmospheric neutrino events to be observed [54] with simple detectors placed in very deep mines is South Africa and in India. Larger samples of up-going muons have been obtained by several detectors: Baksan [55], Frejus [56], MACRO [57], Kamiokande [50] and Super-Kamiokande [52]. The first two detectors have measured fluxes that are compatible with the no-oscillation hypothesis, while Kamiokande, MACRO and Super-Kamiokande have measured deviations from the no-oscillation hypothesis that can be well explained with the existence of oscillations with parameters that are compatible with those indicated by the contained events observations. In the case of up-going muon events one cannot use the two “robust methods” (the $\mu/e$ ratio and the up-down comparison) available for the contained events, since one can observe only $\nu_\mu$ and $\nu_\tau$, in the up-going hemisphere, and therefore the systematic uncertainties are more important. In particular the uncertainty in the prediction of the absolute rate of the fluxes is large of order of 15–20%, however some measurements are affected by smaller uncertainties, and allow to test for the existence of $\nu$ flavor transitions. The most robust prediction is the shape of the zenith angle distribution of the $\phi_{\mu1}$. In the absence of oscillations the shape of the distribution has a minimum (maximum) for the vertical (horizontal) direction(s). This is a consequence of the fact that the decay of the $\pi^\pm$ and $K^\pm$ that generate the neutrinos is more probable for horizontal mesons. This can be understood observing that mesons can decay only during an interval in their trajectory that corresponds to an interaction length; this interval is longer for horizontal mesons that are produced and travel higher in the atmosphere where the air density is lower. This competition between interaction and decay is controlled by geometry and can be predicted accurately. In the presence of oscillations, vertical neutrinos that correspond to longer pathlength can oscillate more, and are more suppressed than horizontal ones and the ratio vertical/horizontal becomes smaller than in the no-oscillation case. This distortion of the predicted shape of the up-going muon flux has been observed by Kamiokande, MACRO (see Fig. 18) and Super-Kamiokande (see Fig. 14.).
Up throughgoing µ flux \(\left(10^{-13} \text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}\right)\)

Fig. 18: The left panel shows a scheme of the MACRO detector, and a scheme of the three classes of atmospheric neutrino events. The right panel shows the measurement of the flux of through-going muons compared with predictions with an without the presence of oscillations.

7.5. **The MACRO data**

The MACRO detector that was located in Hall B of the Gran Sasso Laboratory, under a minimum rock overburden of 3150 kg/cm\(^2\) was a large rectangular box of dimensions 76.6 × 12.3 m\(^3\), divided longitudinally in six supermodules and vertically in a lower part (4.8 m high) and an upper part (4.5 m high) (Fig: 18a). It had three types of detectors which gave redundancy of informations: liquid scintillation counters, limited streamer tubes and nuclear track detectors. This last detector was used only for new particle searches.

The MACRO detector could measure three classes of atmospheric neutrino interactions (see Fig. 18):

- The “Up throughgoing muons” (with \(E_\mu > 1\) GeV) are generated by \(\nu_\mu\) and \(\overline{\nu}_\mu\) interactions in the rock below the detector. The tracking is performed with streamer tubes hits; the time information, provided by scintillation counters, allows the determination of the direction (versus) by the time-of-flight measurement.

- The “Internal Upgoing” (IU) events come from \(\nu_\mu\) interactions inside the lower apparatus. Since two scintillation counters are intercepted, the Time Of Flight method can be applied to identify the upward going muons. The a median parent neutrino energy for these events is \(\sim 4\) GeV. If the atmospheric neutrino anomalies are the results of \(\nu_\mu\) oscillations with maximal mixing and \(10^{-3} < \Delta m^2 < 10^{-2}\) eV\(^2\), one expects a reduction of about a factor of two in the flux of IU events, without any distortion in the shape of the angular distribution.

- The “Up-going Stopping muons” (UGS) are due to \(\nu_\mu\) interactions in the rock below the detector yielding upgoing muon tracks; The “semicontained downgoing muons” (ID) are due to \(\nu_\mu\) induced downgoing tracks with vertex in the lower MACRO. The events are found by means of topological criteria; the lack of time information prevents to distinguish between the two subsamples. An almost equal number of UGS and ID events is expected. In case of oscillations, a similar reduction in the flux of the up stopping events and of the semicontained upgoing muons is expected; no reduction is instead expected for the semicontained downgoing events (which come from neutrinos which travelled \(\sim 20\) km).
A summary of the results of MACRO is given in Fig. 18 and in the Table below:

<table>
<thead>
<tr>
<th>Events</th>
<th>MCno osc</th>
<th>$R = \frac{\text{Data/MCno osc}}{\text{stat} \pm \text{sys} \pm \text{th}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upthroughgoing</td>
<td>809</td>
<td>1122 ± 191 0.721 ± 0.026$<em>{\text{stat}}$ ± 0.043$</em>{\text{sys}}$ ± 0.123$_{\text{th}}$</td>
</tr>
<tr>
<td>IU</td>
<td>154</td>
<td>285 ± 28$<em>{\text{sys}}$ ± 71$</em>{\text{th}}$ 0.54 ± 0.04$<em>{\text{stat}}$ ± 0.05$</em>{\text{sys}}$ ± 0.13$_{\text{th}}$</td>
</tr>
<tr>
<td>ID+UGS</td>
<td>262</td>
<td>375 ± 37$<em>{\text{sys}}$ ± 94$</em>{\text{th}}$ 0.70 ± 0.04$<em>{\text{stat}}$ ± 0.07$</em>{\text{sys}}$ ± 0.17$_{\text{th}}$</td>
</tr>
</tbody>
</table>

The results are consistent with the existence of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations with a best fit point of $\sin^2 2\theta = 1$ and $\Delta m^2 = 2.5 \times 10^{-3}$ eV$^2$ (by coincidence very similar to the best-fit of SK). The data/prediction rates for the three classes of events can be explained qualitatively very simply:

1. The suppression of the rate of internal up-going (IU) events, that are produced by neutrinos with very long pathlength and low energy, can be explained with averaged oscillations, (no distortions of the zenith angle distribution) and the suppression of $\sim 0.5$ corresponds to maximal mixing ($\sin^2 2\theta \sim 1$).

2. The higher energy through-going events have a smaller suppression that is effective mostly for vertical (very long pathlength) events. This can be interpreted as a consequence of the functional dependence of the oscillation probability on the combination $\Delta m^2 L/E_\nu$. These high energy data allow to determine a range of possible $\Delta m^2$ from the shape of the distortion of the zenith angle distribution.

3. The events belonging to the last class (ID + UGS) are a combinations of two sources. The first source (ID) due to internal interactions of down-going neutrinos is little affected by oscillations (because of the short pathlength). The second source (stopping up-going muons) that are suppressed with average oscillations. The combination of the two effects give the average effect $\simeq 0.70$ the.

The region in the plane $(\sin^2 2\theta, \Delta m^2)$ allowed by the MACRO data is shown in Fig. 19, together with the results of other measurements. All results are in reasonable agreement.

![Fig. 19: 90% C.L. allowed region contours for $\nu_{\mu} \leftrightarrow \nu_{\tau}$ oscillations obtained by the Kamiokande, SuperKamiokande, MACRO and Soudan 2 experiments.](image-url)
The two-flavor discussion that we have developed until now corresponds to the special case that corresponds to the condition:

\[ \Delta m_{12}^2 \left( \frac{L}{E} \right)_{\text{max}} \simeq \Delta m_{12}^2 \left( \frac{R_{\oplus}}{0.3 \text{ GeV}} \right) \ll 1, \tag{139} \]

that corresponds to \( \Delta m_{12}^2 \lesssim 10^{-4} \text{ eV}^2 \). In this situation (the “one mass scale dominance” approximation), as discussed in Section 4.4, the oscillation probabilities depend on three parameters:

- One squared mass difference: \( \Delta m^2 \simeq \Delta m_{23}^2 \simeq \Delta m_{13}^2 \).
- Two mixing angles \( \theta_{13} \) and \( \theta_{23} \) that describe the flavor content of the “isolated” (that does not belong to a quasi-degenerate doublet) state that we have labeled \( \nu_3 \):

\[ |\nu_3\rangle = \sin \theta_{13} |\nu_e\rangle = \sin \theta_{13} \cos \theta_{23} |\nu_\mu\rangle = \sin \theta_{13} \sin \theta_{23} |\nu_\tau\rangle. \]

The two-flavor discussion that we have developed until now corresponds to the special case \( \theta_{13} \simeq 0 \) (with the relabeling \( \theta \equiv \theta_{23} \), and the results obtained tell us that the “isolated” state \( \nu_3 \) is a linear combination of \( \nu_\mu \) and \( \nu_\tau \) with approximately equal weight (since \( \sin^2 2\theta_{23} \simeq 1 \) corresponds to \( \theta_{23} = 45^\circ \)).

It is clearly of considerable interest to study both angles \( \theta_{23} \) and \( \theta_{13} \). It turns out that the best measurement of \( \theta_{13} \) has been obtained not with atmospheric neutrinos but with neutrinos from reactors (see Section 9.), and gives an upper limit \( \sin^2 \theta_{13} < 0.05 \). Therefore “a posteriori” the two-flavor analysis of atmospheric neutrinos turns out to be a surprisingly good approximation. Analysis of the atmospheric neutrino data in terms of three flavor oscillations in the framework of the one mass scale dominance approximation have been performed by SK. The result (see Fig. 33) gives allowed intervals for \( \Delta m^2 \) and \( \theta_{23} \) that are very close to the intervals obtained in the two-flavor analysis, while for the other angle \( \theta_{13} \) one obtains only an upper limit \( \sin^2 \theta_{13} \leq 0.75 \). The upper limit for the angle \( \theta_{13} \) may seem surprisingly large considering that these angle controls the oscillations of electron neutrinos, and all \( \mu \)-like events are compatible with the absence of oscillations. The poor limit if \( \theta_{13} \) can be understood as the consequence of an interesting cancellation effect. After inclusion of oscillations, the flux of \( e \)-like events can be schematically written as:

\[ e = e_0 \left[ 1 - P_{\nu_e \rightarrow \nu_\mu} - P_{\nu_e \rightarrow \nu_\tau} \right] + \mu_0 P_{\nu_\mu \rightarrow \nu_e} \tag{140} \]

in the limit of small \( \Delta m_{12}^2 \) one has the relations:

\[
\begin{align*}
P_{\nu_e \rightarrow \nu_\mu} &= P_{\nu_\mu \rightarrow \nu_e}, \\
P_{\nu_e \rightarrow \nu_\mu} \sin^2 \theta_{23} &= P_{\nu_e \rightarrow \nu_\tau} \cos^2 \theta_{23}.
\end{align*}
\tag{141}
\]

Using the result \( \sin^2 \theta_{23} \simeq \cos^2 \theta_{23} \simeq 0.5 \) (that follows from \( \sin^2 2\theta_{23} \simeq 1 \)) one obtains: \( P_{\nu_\mu \rightarrow \nu_e} = P_{\nu_e \rightarrow \nu_\mu} \simeq P_{\nu_e \rightarrow \nu_\tau} \simeq P \), and substituting into (140) one obtains:

\[ e \simeq e_0 \left[ 1 - 2 P \right] + \mu_0 P \simeq e_0 \left\{ \left[ 1 - 2 P \right] + \left( \frac{\mu_0}{e_0} \right) P \right\} \simeq e_0, \tag{142} \]
where we have used the result that before oscillations $\mu_0/e_0 \simeq 2$. Therefore even if the oscillations of
electron neutrinos are rather large, the effect on the flux is small because one has a disappearance effect
that has two channels ($\nu_e \rightarrow \nu_\mu$ and $\nu_e \rightarrow \nu_\tau$) with approximately equal probability, and one appearance
effect ($\nu_\mu \rightarrow \nu_e$ with the same transition probability, but a source that is twice as large, and to first order
the two effects cancel.

7.7. The $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$ hypothesis

Since the effect observed in atmospheric neutrinos is the disappearance of $\mu$-like events, it is natural
to ask the question if the physical process observed is in fact due to transitions $\nu_\mu \rightarrow \nu_\tau$, or if it is
instead due to transitions $\nu_\mu \rightarrow \nu_{\text{sterile}}$, where $\nu_{\text{sterile}}$ is a singlet under the $SU(2) \otimes U(1)$ group of the
electroweak interactions that is a particle that does not interact with the $W$ and $Z$ bosons and is therefore
completely invisible.

The two type of oscillations can be distinguished using three methods:

(i) The most straightforward method is of course the detection of $\tau^\pm$ leptons, that can be produced by
the charged current interactions of $\nu_\tau$ and $\bar{\nu}_\tau$ in the case of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations, but are absent in
the case of $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$. However the expected rate of production of $\tau^\pm$ is expected to be small
because only high energy neutrinos above threshold can contribute, and the identification of these
events is experimentally very difficult.

(ii) $\nu_\tau$ and $\bar{\nu}_\tau$ can interact with neutral current interactions, while a sterile neutrino is by definition
non interacting; therefore in the hypothesis of $\nu_\mu \rightarrow \nu_{\text{sterile}}$ oscillations, one should also observe a
suppression of the rate of neutral current interactions.

(iii) Another method to distinguish $\nu_\mu \leftrightarrow \nu_\tau$ and $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$ oscillations is the fact that in this second
hypothesis the oscillation probability is modified by matter effects. As discussed before matter
effects are absent in the case of $\nu_\mu \leftrightarrow \nu_\tau$ oscillations since the effective potentials for $\nu_\mu$ and $\nu_\tau$ are equal:

\[ V(\nu_\mu) = V(\nu_\tau) = \sqrt{2} G_F \left[ \frac{N_p}{2} - \frac{N_e}{2} - \frac{N_n}{2} \right] = -\sqrt{2} G_F \frac{N_n}{2}, \quad (143) \]

where $N_p$, $N_e$ and $N_n$ are the densities of protons, electrons and neutrons, and we have used the
condition of electric neutrality ($N_e = N_p$). On the other hand, the absence of interactions results
in $V(\nu_{\text{sterile}}) = 0$, with a potential difference:

\[ V_{\mu \tau} = -\sqrt{2} G_F \frac{N_n}{2} \simeq -10^{-13} \text{ eV} \left( \frac{\rho}{5 \text{ g cm}^{-3}} \right). \quad (144) \]

Note that the potential difference $V_{\mu \tau}$ differs by a factor of 2 and by a sign from the potential
difference $V_{\nu \tau} = V_{e \tau}$.

The potential difference becomes significant when it is comparable or larger than the energy dif-
fERENCE between different neutrinos eigenmass states:

\[ |V_{\mu \tau}| \geq |E_2 - E_1| \simeq \frac{\Delta m^2}{2E_\nu}, \quad (145) \]

that is for a neutrino energy larger than: $E_\nu \geq |\Delta m^2|/(2|V_{\mu \tau}|) \sim 60$ GeV, where we have used $|\Delta m^2| = 3 \times 10^{-3}$ eV$^2$ and a density $\sim 5$ g cm$^{-3}$ for the numerical estimate. It follows that the matter effects are
a negligible effect for the sub-GeV and multi-GeV events and become important (and then dominant)
only at higher energy. In this case the most important consequence of the matter effects is a suppression
of the oscillation probability at large energy. This suppression can be understood from the expression for
the effective mixing parameter in matter $\sin^2 2\theta_m$, that for maximal mixing in vacuum is:

$$\sin^2 2\theta_m = \frac{1}{1 + (2 V_{\mu\mu}^2 E_\nu/\Delta m^2)^2},$$  \hspace{1cm} (146)$$

that is suppressed $\propto E_\nu^{-2}$ at large neutrino energies.

Super-Kamiokande has analysed its data to test if the $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$ oscillation hypothesis can provide a valid answer. In order to test this hypothesis the collaboration has analysed three samples of data:

(i) A neutral current (NC) enriched sample constructed as a set of contained events where the highest energy particle (ring) is $e$-like. This sample of events contains only a small contamination of $\nu_\mu$ ($\overline{\nu}_\mu$) CC interactions, and is composed by a combination of $\nu_e$ ($\overline{\nu}_e$) CC interactions and NC (neutral current) interactions.

(ii) A sample of large energy, $\mu$-like, contained events selected with large visible energy: $E_{\text{vis}} > 5$ GeV.

(iii) The through-going upward-going muon events.

The first type of events allows to study if neutral current interactions are suppressed (oscillations with $\nu_{\text{sterile}}$) or not (oscillations with $\nu_\mu$). In the first case one should observe an up-down asymmetry of the events due to the disappearance of up-going neutral current interactions. The second and third classes of events select high energy neutrinos. The oscillation of these high energy neutrinos is suppressed in the $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$ case, because of the existence of matter effects. The data (see Fig. 20) shows no evidence for an up-down asymmetry of the neutral current events, and no evidence for a suppression of the mixing parameter $\sin^2 2\theta$ with energy, and this has allowed the SK collaboration [53] to eliminate the $\nu_\mu \leftrightarrow \nu_{\text{sterile}}$ hypothesis at more than 3 sigmas. This results is confirmed by the analysis of throughgoing muons of the MACRO collaboration [59].

It is natural to consider a more general situation where muon (anti-)neutrinos oscillate into a linear combination of type $\cos \xi \nu_\mu + \sin \xi \nu_{\text{sterile}}$. A preliminary analysis of SK results in a 90% C.L. upper limit $\sin^2 \xi \leq 0.25$ (see then bottom panel of Fig. 21).
Fig. 21: Limit on the Sterile Content of Atmospheric Neutrino Oscillations. The limit on the sterile content is $\sin^2 \xi < 0.25$ at 90% C.L.

7.8. Alternative explanations

Several alternative mechanisms have been proposed as explanations of the atmospheric neutrino data, based for example on the possibility of neutrino decay [62], violations of the equivalence principle [63,64], the existence of Flavor Changing Neutral Currents (FCNC) [65], or the effect of large extra dimensions [66–68]. Most of these models are excluded by the data (for reviews see [60,61]). The reason for this is that atmospheric neutrino data probe three decades in pathlength $L$ and four decades in energy $E$. Such a wide dynamical range severely constrains deviations from the standard $L/E$ behavior of the $P_{\mu\tau}$ transition probability. Therefore, $P_{\mu\tau}$ seems to be (dominantly) a function of $L/E$. Models that still give a reasonable fit are those that have a probability that is a function of $L/E$ but with a non-oscillatory behaviour, such as in neutrino decay ([62]), or in the decoherence scenario of [69]. In these models the survival probability (after averaging) has the same “asymptotic” form for large $L/E$, but a monotonic behaviour. The large energy-angle smearing of SK prevents a clear discrimination (see Fig. 22).

Fig. 22: Survival probability $P_{\nu_{\mu}\rightarrow\nu_{\mu}}$ capable of explaining the atmospheric neutrino data. The oscillating curve is the result of a standard oscillation scenario (with best fit parameters). The solid curve can be obtained in a neutrino decay scenario [62].


8. SOLAR NEUTRINOS

8.1. The Solar neutrino fluxes

The energy of the sun is produced in nuclear fusion reactions (for reviews see [70,71]). The process that generates the energy is the combination of 4 protons and 2 electrons to produce a helium-4 nucleus and two neutrinos:

\[ 4p + 2e^- \rightarrow ^4\text{He} + 2\nu_e. \]  

(147)

This reaction can happen with different sets (or “cycles” of reactions) that produce the same final particles, but that result in different energy distributions for the neutrinos, therefore for a detailed prediction of the neutrino fluxes one needs to build a model of the sun and compute the contributions of the different cycles.

Main sequence stars burn their hydrogen into helium with three \(pp\) cycles (see Fig. 23) and/or with a CNO bi-cycle. The \(pp\) cycles are the dominant mechanism for energy production in cooler (lower mass) stars while the CNO bi-cycle mechanism is dominant for hotter (larger mass) stars. In the sun \(98.5\%\) of the energy is produced with the \(pp\) cycles and only \(1.5\%\) with the CNO bi-cycle (that to a good approximation can be neglected).
The main components of the solar neutrino flux are shown in Fig. 24, and are also tabulated in Table 1. The different components are the result of different cycles of nuclear burning that all result in the same effect of combining 4 protons and two electrons to produce a helium-4 and two neutrinos, but produce neutrino with different energy distributions.

A brief description of the \( pp \) cycles can be summarized as follows. The first step in the \( pp \) cycles is the formation of deuterium \((^2H)\). This happen most of the time with the weak interaction two-body reaction:

\[
pp \rightarrow ^2H + e^+ + \nu_e \quad [pp \text{ neutrinos}].
\]

This reaction is the source of the so called \( pp \) neutrinos. In the dense environment of the solar core also the three body reaction

\[
p + e^- + p \rightarrow ^2H + \nu_e \quad [pep \text{ neutrinos}]
\]
Table 1: Predictions of the solar neutrino fluxes.

<table>
<thead>
<tr>
<th>Source</th>
<th>Reaction</th>
<th>$\langle E_\nu \rangle _\odot$ (MeV)</th>
<th>Flux (cm$^{-2}$sec)$^{-1}$</th>
<th>Cl (SNU)</th>
<th>Ga (SNU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pp</td>
<td>$p + p \rightarrow , , ^2\text{H} + e^+ + \nu_e$</td>
<td>0.2668</td>
<td>$5.95 \times 10^{10}$</td>
<td>0.0</td>
<td>69.7</td>
</tr>
<tr>
<td>pep</td>
<td>$p + e^- + p \rightarrow , , ^2\text{H} + \nu_e$</td>
<td>1.445</td>
<td>$1.40 \times 10^8$</td>
<td>0.22</td>
<td>2.8</td>
</tr>
<tr>
<td>hep</td>
<td>$^3\text{He} + p \rightarrow , , ^4\text{He} + e^+\nu_e$</td>
<td>9.628</td>
<td>$9.30 \times 10^3$</td>
<td>0.04</td>
<td>0.1</td>
</tr>
<tr>
<td>$^7\text{Be}$</td>
<td>$^7\text{Be} + e^- \rightarrow , , ^7\text{Li} + \nu_e$</td>
<td>0.814</td>
<td>$4.77 \times 10^9$</td>
<td>1.15</td>
<td>34.2</td>
</tr>
<tr>
<td>$^8\text{B}$</td>
<td>$^8\text{B} \rightarrow , , ^8\text{Be} + e^+\nu_e$</td>
<td>6.735</td>
<td>$5.05 \times 10^6$</td>
<td>5.76</td>
<td>12.1</td>
</tr>
<tr>
<td>$^{13}\text{N}$</td>
<td>$^{13}\text{N} \rightarrow , , ^{13}\text{C} + e^+\nu_e$</td>
<td>0.706</td>
<td>$5.48 \times 10^8$</td>
<td>0.09</td>
<td>3.4</td>
</tr>
<tr>
<td>$^{15}\text{O}$</td>
<td>$^{15}\text{O} \rightarrow , , ^{15}\text{N} + e^+\nu_e$</td>
<td>0.996</td>
<td>$4.80 \times 10^8$</td>
<td>0.33</td>
<td>5.5</td>
</tr>
<tr>
<td>$^{17}\text{F}$</td>
<td>$^{17}\text{F} \rightarrow , , ^{17}\text{O} + e^+\nu_e$</td>
<td>0.999</td>
<td>$5.63 \times 10^6$</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$7.6 , ^{+1.3}_{-1.1}$</td>
<td>$128 , ^{+6.5}_{-6.0}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

can happen and results in an approximately monochromatic neutrino line ($pep$ neutrinos with $E_\nu = 1.41$ MeV). The deuterium generated in these reactions rapidly absorbs one proton with the reaction $d + p \rightarrow \, \, ^3\text{He} + \gamma$, producing Helium-3. The $^3\text{He}$ is consumed in interactions with another $^3\text{He}$ or with a $^4\text{He}$ nucleus. In the first case (the 3–3 branch) the reaction is: $^3\text{He} + ^3\text{He} \rightarrow \, \, ^4\text{He} + \gamma$ that completes the cycle ($pp$–I chain) with the formation of helium-4. In the other (3–4) branch one has the reaction $^3\text{He} + ^4\text{He} \rightarrow \, \, ^7\text{Be} + \gamma$ producing beryllium-7. The two possible fates of the beryllium nuclei determine an additional branching. The most likely event is an electron capture reaction:

$$^7\text{Be} + e^- \rightarrow \, \, ^7\text{Li} + \nu_e \quad [^7\text{Be} \, \, \text{neutrinos}].$$

The lithium nucleus is formed in the ground state or in an excited one, and therefore the $^7\text{Be}$ neutrinos form two lines with 0.8631 MeV (89.7%) and 0.3855 MeV (10.3%). The Lithium then absorbs a proton to produce two $^4\text{He}$ nuclei completing the ($pp$–II) cycle.

The second branch for the Beryllium nuclei is the absorption of a proton to produce $^8\text{B}$oron, that undergoes a beta decay into two $^4\text{He}$ nuclei:

$$^8\text{B} \rightarrow \, \, 2^4\text{He} + e^+ + \nu_e \quad [^8\text{B} \, \, \text{neutrinos}],$$

completing the ($pp$–III) cycle. The $^8\text{B}$ neutrinos extend up to the highest energy of 14 MeV, and even if they are only a fraction $10^{-4}$ of the flux, play a very important role in the detection of the solar neutrinos. It is important to understand that the kinetic energy of the interacting particles is of order $E_{\text{kin}} \sim T_{\text{center}} \approx 1.5 \times 10^7$ Kelvin $\sim 1.5$ KeV is small with respect to the energy released in the reaction, therefore the shape of the neutrino spectra emitted in each reaction is determined uniquely by particle physics, with no dependence from the modeling of the sun that has simply the role of determining the relative importance of the different branches.
8.2. Detection of Solar Neutrinos

Measurements of the solar neutrino fluxes have been reported from several experiments using four different methods of detection. They all have reported measurements smaller than the prediction.

**The Chlorine experiment**

In the “Chlorine experiment” one exposes $^{37}$Cl to solar neutrinos, and detect the argon-37 (a radioactive unstable nucleus) produced in the reaction

$$\nu_e + ^{37}\text{Cl} \rightarrow ^{37}\text{Ar} + e^{-}.$$  

(150)

This reaction has a threshold energy of 0.814 MeV. An experiment based on this technique, has been constructed by Raymond Davis and his collaborators in the Homestake mine in South Dakota (US), and has been taking data from 1969 to 1993. This detector first reported a deficit in the neutrino flux [75].

**The Gallium experiment**

A lower energy threshold ($E = 0.214$ MeV) can be obtained using a Gallium target, and the reaction

$$\nu_e + ^{71}\text{Ga} \rightarrow ^{71}\text{Ge} + e^{-}.$$  

(151)

The Chlorine and Gallium experiments are known as **Radio–chemical methods**, they both are sensitive only to electron neutrinos, in fact since an $e^-$ is present in the final state $\nu_\mu$, $\nu_\tau$ and anti-neutrinos cannot induce the nuclear transitions (150) and (151). Two detectors based on the gallium technique have been taking data: GALLEX [76] in the Gran Sasso underground laboratory and SAGE [77] in Russia.

The quantity that is measurable in the radiochemical experiments such as the Chlorine and Gallium experiments is a capture rate, that is the probability per unit time that a target nucleus exposed to the solar neutrinos “captures” a solar neutrino with a reaction of the type (150) or (151). The capture rate can be calculated as:

$$C_j = \int dE \, \phi_{\nu_{\odot}}(E) \, \sigma_j(E),$$  

(152)

where $\phi_{\nu_{\odot}}(E)$ is the flux of solar neutrinos with energy $E$ and $\sigma_j(E)$ is the cross section (the subscript $j$ indicates the reaction considered). The capture rate is measured in units of sec$^{-1}$. It has become customary to use the unit:

$$1 \text{ SNU} \equiv 1 \text{ Solar Neutrino Unit} = 10^{-36} \text{ sec}^{-1}.$$  

(153)

In Table 1 we show the estimates for the capture rates in chlorine and gallium, and the contributions of the different components of solar neutrinos. Note that the Chlorine experiment is mostly sensitive to the intensity of the boron and beryllium neutrinos, while the most abundant $pp$ neutrinos are below threshold, while more than half the rate for Gallium experiments is predicted to be generated by $pp$ neutrinos.

**Electron Scattering**

Solar neutrinos are also detectable observing the Elastic Scattering on electrons:

$$\nu_x + e^- \rightarrow \nu_x + e^-.$$  

(154)

This reaction does not have a threshold, however it is detectable above the natural radioactivity background only when the final state electron has a sufficiently high energy. Threshold as low as $E > 5$ MeV, have been obtained. This method can only reveal the highest energy ($^8\text{B}$) solar neutrinos. The electron scattering method is sensitive to all neutrino types, however the cross section for $\nu_x$ is is approximately 7 larger than for $\nu_\mu$ and $\nu_\tau$. The final state electron in the reaction (154) is emitted in a direction strongly correlated with the neutrino one. This technique has been first used by the Kamiokande [79] detector in Japan, and then by the larger size Super-Kamiokande [83] detector. The cross section for neutrino
electron-scattering is given by the expression:

\[
\frac{d\sigma_{\nu_e}}{dT} = \frac{2G_F^2 m_e^2}{\pi} \left[ g_L^2 + g_R^2 \left( 1 - \frac{T}{E_{\nu}} \right)^2 - g_L g_R \frac{m_e c^2 T}{E_{\nu}} \right],
\]

(155)

where \( T \) is the kinetic energy of the final state electron and the constants \( g_L^2 \) and \( g_R^2 \) have values:

\[
g_L^2 = \begin{cases} \left( \frac{1}{2} + \sin^2 \theta_W \right)^2 & \approx 0.536, \quad \nu_e \\ \sin^4 \theta_W & \approx 0.0538, \quad \bar{\nu}_e \\ \left( -\frac{1}{2} + \sin^2 \theta_W \right)^2 & \approx 0.0719, \quad \nu_i \\ \sin^4 \theta_W & \approx 0.0538, \quad \bar{\nu}_i \end{cases}
\]

\[
g_R^2 = \begin{cases} \sin^4 \theta_W & \approx 0.0538, \quad \nu_e \\ \left( \frac{1}{2} + \sin^2 \theta_W \right)^2 & \approx 0.536, \quad \bar{\nu}_e \\ \sin^4 \theta_W & \approx 0.0538, \quad \nu_i \\ \left( -\frac{1}{2} + \sin^2 \theta_W \right)^2 & \approx 0.0719, \quad \bar{\nu}_i \end{cases}
\]

For electron neutrinos the cross section is approximately \( \sigma(\nu_e) \approx 1.0 \times 10^{-44} \text{ E}_{\text{MeV}} \text{ cm}^2 \) while for \( \nu_\mu \) and \( \nu_\tau \) the cross section is \( \sigma(\nu_{\mu,\tau}) \approx 0.15 \times 10^{-44} \text{ E}_{\text{MeV}} \text{ cm}^2 \) is approximately 7 times smaller.

**Heavy Water**

Two other methods to measure solar neutrinos make use of a deuterium target. The Sudbury Neutrino Observatory (SNO) detector in Canada [85] contains 1 Kton of heavy water. One could expect that the reaction \( \nu_e + n \to e^- + p \) plays an important role in the detection of solar neutrinos, however free neutrons do not exist in nature, therefore this reaction can only be observed with bound neutrons in nuclei, and for nearly all nuclei the threshold for the reaction \( \nu_e + (A;Z) \to e^- + (A;Z + 1) \) is too high for solar neutrinos. The deuterium nucleus with its small binding energy (2.2 MeV) is the closest approximation of free neutrons, and the reaction

\[
\nu_e + d \to e^- + p + p
\]

(156)

has an energy threshold of only 1.44 MeV. The positron detected in the reaction can be observed. The reaction (156) is only sensitive to \( \nu_e \).

Deuterium allows also to detect solar neutrinos with the neutral-current reaction:

\[
\nu_x + d \to \nu_x + p + n,
\]

(157)

where deuterium in broken up into its nucleon constituents. This reaction has a threshold of 2.2 MeV, and a cross section that is identical for all flavors; it can be detected observing the photons emitted in the capture of the final state neutron: \( n + (A;Z) \to (A + 1;Z) + \gamma \).

**8.3. Experimental Results**

In Fig. 25 is shown the distribution of the angle \( \theta_{\text{sun}} \) (angle with respect to the the sun direction) for all particles detected particles in Super-Kamiokande with an energy larger than 5 MeV. There is clear evidence for a contribution due to solar neutrinos above a flat background. With the electron scattering technique the sun is “seen” in neutrinos in the sense that a real image (even if rather blurry) can be constructed using neutrinos.
Fig. 25: Distribution of \( \cos \theta_{\odot} \) of Super-Kamiokande events. The peak at \( \cos \theta_{\odot} \approx 1 \) is the contribution of events generated by electron scattering: \( \nu_e + e^- \rightarrow \nu_e + e^- \), superimposed on a flat background.

The energy and zenith angle distribution of the SK events are shown in Fig. 26. It should be noted that there are not statistically significant distortions with respect to a no-oscillation prediction. This sets important constraints on possible solutions in terms of \( \nu \) oscillations.

Fig. 26: The left panel shows the energy spectrum of the events measured by Super-Kamiokande in the solar neutrino signal plotted as the ratio to the Standard Solar Model calculation of [72]. The horizontal solid line is the ratio for the total flux, the dotted band around this line is the systematic uncertainty due to the energy scale. There is no evidence for a spectral distortion. A flat ratio gives \( \chi^2 / \text{d.o.f.} = 19.0 / 18 \). The right panel shows the zenith angle distribution of the Super-Kamiokande events in the solar neutrino signal. Neutrinos detected during the night have traveled inside the Earth. From [83].

The Sudbury Neutrino Observatory (SNO) in Canada uses as sensitive material one kiloton of heavy water (D\(_2\)O) contained in a transparent acrylic shell of 12 meter diameter. The Cherenkov photons emitted by charged particles are detected by a system of 9546 PMT’s mounted on a geodesic structure of 17.8 meters of diameter. The geodesic structure is also immersed in ultra-pure water to provide shielding. The right panel of Fig. 27 shows results obtained with the SNO detector. The top panel shows the distribution in \( \cos \theta_{\odot} \) of all events with energy larger than 5 MeV. For a heavy water detector one has three different contributions:

- Events due to electron scattering \( (\nu_e + e^- \rightarrow \nu_e + e^-) \) that have an angular distribution peaked at small \( \theta_{\odot} \).
- Events due to the charged current scattering process \( \nu_e + d \rightarrow e^- + p + p \). In this process the electron is produced with an energy \( E_e \approx E_\nu - 1.44 \) MeV and an approximately flat angular distribution with a small negative slope in \( \cos \theta_{\odot} \), so that the most likely scattering is backward.
- Finally the 6.5 MeV photons produced in the neutron capture process \( n + d \rightarrow T + \gamma \) give an isotropic (flat in \( \cos \theta_{\odot} \)) distribution. Most neutrons are produced by the neutral current reaction \( \nu_x + d \rightarrow \nu_x + p + n \)
One can see that because of the different angular, and energy (see the bottom panel in Fig. 27) it is possible to disentangle the three contributions.

The central panel in Fig. 27 shows that in the inner fiducial volume the event rate is uniform, as expected for neutrino interactions (while background events induced by radioactivity is concentrated near the outer region).

![Fig. 27: Left panel: the SNO detector. The inner vessel contains 1 Kton of ultra-pure heavy water. Right panels: (a) Distribution of \( \cos \theta_C \) of the detected events in the inner fiducial volume (\( R < 0.55 \) m). (b) Distribution of the volume weighted variable. (c) Kinetic energy distribution of the events in the inner fiducial volume. Also shown are the Monte Carlo prediction for the three classes of events: CC (\( \nu_e + d \rightarrow e^+ + p + n \)), Electron Scattering (ES: \( \nu_x + e^- \rightarrow \nu_x + e^- \)), and Neutral currents (\( \nu_x + d \rightarrow p + n \)). The shapes of the curves is fixed, but the relative normalization is fitted to the data. The bad indicated \( \pm 1 \sigma \) uncertainties. The dashed lines are the sum of all contributions. All distributions are for a detected kinetic energy larger than 5 MeV. From [86].

In Table 2 we show the results of the different measurements of solar neutrinos and compare with the predictions based on the solar model of Bahcall et al [72]. The last column in the Table shows the ratio of the measured rate with the theoretical prediction based on the no-oscillation hypothesis. All experiments detect a significant deficit.

The top part of Table 2 gives the results for the radiochemical experiments: the chlorine one [75], and the two Gallium experiments GALLEX [76,78] and SAGE [77]. For this part of the Table the unit used for both data and prediction is SNU (or \( 10^{-36} \) neutrino captures per nucleus and per second). The bottom part of the Table gives results of the measurement of the highest energy part of the solar neutrino flux, where only neutrinos from Boron-8 decay contribute, and the unit used for both measurement and prediction is a flux: \( 10^6 \) (cm\(^2\) s\(^{-1}\)). This mean that the results are given in term of an undistorted flux of neutrinos with only the absolute normalization free. The flux of \( ^8 \)B neutrinos has been measured with three different techniques: (i) electron elastic scattering (ES), (ii) CC \( \nu_e \)-deuterium scattering \( \nu_e + d \rightarrow e^- pp \), and (iii) the observations of the capture of neutrons produced in the NC reactions \( \nu_x + d \rightarrow \nu_x + p + n \). The electron scattering experiment has been performed by three experiments...
Table 2: Measurements of solar neutrinos and comparison with no-oscillations predictions.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Prediction</th>
<th>Data</th>
<th>Data/Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorine</td>
<td>7.6$^{+1.3}_{-1.1}$</td>
<td>2.56 ± 0.23</td>
<td>0.34 ± 0.06</td>
</tr>
<tr>
<td>GALLEX + GNO</td>
<td>128$^{+6.7}_{-7.8}$</td>
<td>74.1$^{+0.09}_{-0.07}$</td>
<td>0.58 ± 0.07</td>
</tr>
<tr>
<td>SAGE</td>
<td>128$^{+9}_{-7}$</td>
<td>75.4$^{+7.8}_{-7.4}$</td>
<td>0.59 ± 0.07</td>
</tr>
<tr>
<td>$^8$B–ES [Kamiokande]</td>
<td>5.05$^{+1.0}_{-0.8}$</td>
<td>2.80 ± 0.40</td>
<td>0.55 ± 0.13</td>
</tr>
<tr>
<td>$^8$B–ES [SuperKamiokande]</td>
<td>5.05$^{+1.0}_{-0.8}$</td>
<td>2.40$^{+0.09}_{-0.07}$</td>
<td>0.48 ± 0.09</td>
</tr>
<tr>
<td>$^8$B–ES [SNO]</td>
<td>5.05$^{+1.0}_{-0.8}$</td>
<td>2.39 ± 0.24</td>
<td>0.47 ± 0.12</td>
</tr>
<tr>
<td>$^8$B–CC [SNO]</td>
<td>5.05$^{+1.0}_{-0.8}$</td>
<td>1.76 ± 0.06</td>
<td>0.35 ± 0.08</td>
</tr>
<tr>
<td>$^8$B–NC [SNO]</td>
<td>5.05$^{+1.0}_{-0.8}$</td>
<td>5.09 ± 0.44</td>
<td>1.0 ± 0.20</td>
</tr>
</tbody>
</table>

(Kamiokande [82], Super-Kamiokande [83] and SNO [85] the measurements using deuterium have been performed by SNO alone.

Inspecting Table 2 and comparing the measurements to the predictions one can make several important points:

1. There is a statistical very significant deficit of neutrinos that is observed by all measurements, except for the measurement that involves neutral currents. This is evidence for the “disappearance” of $\nu_e$.

2. The deficit is not equal for different experiments that probe different energy regions of the spectrum. The Gallium experiments, that probes the lowest energy solar neutrinos, observes the smallest deficit, while the chlorine experiment observes the largest deficit. This is evidence for a non trivial energy dependence of the disappearance probability $P(\nu_e \rightarrow \nu_x)$.

3. The three measurements of the highest energy ($^8$B) neutrinos with the electron-scattering technique are all in good agreement with each other, however the measurement of SNO using CC interactions on deuterium indicates a larger deficit. A careful statistical analysis [85] shows that the measurements of the ES and CC method differ by 3.3 standard deviations. This can be explained assuming that the “disappearing” $\nu_e$ are transformed into $\nu_\mu$ or $\nu_\tau$ that can still contribute a suppressed but non zero contribution to the electron scattering event rate (see Fig. 28).

4. This interpretation is confirmed by the agreement of prediction and observations for the neutral current events.

The “real-time” experiments of Super-Kamiokande and SNO, also give a time distribution and an energy spectrum of the detected neutrino events. This allows also to perform more detailed studies, searching for effects such as:

1. A deformation of the energy spectrum of the neutrinos with respect to a prediction that depend uniquely on particle physics. No significant deformation has been observed (see Fig. 26).

2. A difference of the detected rate between day and night (or more in general on the orientation of the sun with respect to the detector, that is on the sun zenith angle). Neutrinos detected during the night have traversed the Earth, and it is possible that the presence of the Earth matter modifies the flavor transition probabilities.
8.4. Interpretation

In the presence of neutrino flavor transitions the event rates for the solar neutrino detectors are suppressed because the electron neutrino flux at the Earth becomes

$$\phi_{\nu_{e}\odot}(E) = \phi_{\nu_{e}}^{\text{SSM}}(E) \times P_{\nu_{e}\rightarrow\nu_{e}}(E),$$

(158)

where $\phi_{\nu_{e}}^{\text{SSM}}(E)$ is the Standard Solar Model (SSM) prediction. In the case of electron scattering, the effective flux that can generate the scattering is a linear combination (with different weights) of the $\nu_{e}$ and $\nu_{\mu,\tau}$ fluxes that reach the earth. Therefore one has to make the substitution:

$$\phi_{\nu_{e}}^{\text{SSM}}(E) \rightarrow \phi_{\nu_{e}}^{\text{SSM}}(E) \times \left\{ P_{\nu_{e}\rightarrow\nu_{e}}(E) + [1 - P_{\nu_{e}\rightarrow\nu_{e}}(E)] \times \frac{\sigma_{\nu_{e}\rightarrow\nu_{e}}(E)}{\sigma_{\nu_{e}}(E)} \right\}. \quad (159)$$

Qualitatively it is clear that neutrino oscillations is a natural candidate for explaining the data. The question is now to verify if it is possible to find values of the oscillation parameters (squared mass differences and mixing angles) that give predictions that are in quantitative agreement with the experimental data. For simplicity we will consider the case of two flavor ($\nu_{e}$--$\nu_{\mu}$ or $\nu_{\mu}$--$\nu_{\tau}$) mixing. In this case the $\nu_{e}$ survival probability, averaged over all creation points of the neutrinos, will be a function of the two parameters $\theta$ and $\Delta m_{21}^2$, where $\theta \in [0, \frac{\pi}{2}]$ is the 2-flavor mixing angle, and $\Delta m_{21}^2$ the squared mass difference. This description of the oscillation of solar $\nu_{e}$ in terms of simple two-flavor mixing is actually a very good approximation of the most general case in the light of the results of atmospheric and reactor neutrinos. These results indicate that $\nu_{e}$’s do not participate in the “fast” oscillations related to the squared mass difference $\Delta m_{21}^2$. This can be interpreted a rather stringent upper limit on the size of $\theta_{13}$. It is easy to demonstrate that in the limit of $\theta_{13} \rightarrow 0$ the oscillation probability $P_{\nu_{e}\rightarrow\nu_{e}}$ is identical to what is calculated with two-flavor mixing with the substitutions $\Delta m_{21}^2 \rightarrow \Delta m_{12}^2$ and $\theta \rightarrow \theta_{12}$.
Fig. 29: Oscillation Predictions of SK Spectral Distortion (Left) and Day/Night Asymmetry (Right). The spectral distortion is depicted in contours of equal $\chi^2$. The shown values of both the $\chi^2$ and the day/night asymmetry contours increase logarithmically. Superimposed in gray are the usual solution areas. From [84]

A detailed study of the parameter space (that is the plane $(\tan^2 \theta_{12}, \Delta m^2_{12})$) reveals that there are values of the parameters that allow to reproduce the total rates measured by the different experiments. These disconnected regions are shown in Fig. 29 as dark-shaded (red) regions. The 4 disconnected parts are commonly named:

(a) **LMA**

The Large Mixing Angle solution corresponds to the region with small mixing angle $\tan^2 \theta \simeq 10^{-3}$ and $\Delta m^2 \in [0.9, 10] \times 10^{-5}$ eV$^2$.

(b) **SMA**

The Small Mixing Angle solution corresponds to the region with large mixing angle $\tan^2 \theta \in [0.2, 0.5]$ and $\Delta m^2 \in [0.4, 1] \times 10^{-5}$ eV$^2$.

(c) **LOW**

The so called “Low” solution corresponds to a region again with large mixing angle $\tan^2 \theta \simeq 0.5$ but with lower $\Delta m^2$: $\Delta m^2 \in [0.8, 1] \times 10^{-7}$ eV$^2$.

(d) **Vacuum**

The “vacuum solution” corresponds to a subset of disconnected region with $\tan^2 \theta$ around one (that is $\theta \simeq 45^\circ$, and $\Delta m^2 \sim 10^{-10}$ eV$^2$).

A detailed analysis of the data of Super-Kamiokande and SNO (including the spectral shape and the day–night differences) disfavors some of these solutions, and the LMA solution emerges as the most likely solution. This is indicated in Fig. 29 and Fig. 30.

The complexity of the shape of the parameter space region that can explain the detected rates of the solar neutrino experiments, is due to some remarkable features of the flavor transition probabilities in matter with a varying density that are discussed in the following.

It is simple to see that because of the shape of the oscillation probability in matter for fixed neutrino

\footnote{Note that the since the mixing angle $\theta_{12}$ is defined in the interval $\theta \in [0, \pi]$, $\sin^2 \theta$ is not a good variable because two physically different points ($\theta$ and $\pi - \theta$) are mapped into the same value. On the other hand $\sin^2 \theta_{12}$ or $\tan^2 \theta_{12}$ are good variables.}
Fig. 30: Allowed regions in the parameter space \( \Delta m_{12}^2, \tan^2 \theta_{12} \) (assuming \( \theta_{13} \approx 0 \)) including all measurements of solar neutrinos. The best fit point is: \( \Delta m_{12}^2 = 5 \times 10^{-5} \text{ eV}^2 \), and \( \tan^2 \theta_{12} = 0.32 \). From [87].

Energy \( E_\nu \), the set of points in the parameter space \( (\sin^2 2\theta, \Delta m^2) \) that give a desired survival probability \( P_{\nu_e \rightarrow \nu_e} = \tilde{P} \) has the shape of a “triangle” (see Fig. 31).

The allowed region in parameter space has recently been very much reduced thanks to the data of SNO. The results are shown in Fig. 30 (from [87]). The top panel shows the allowed region obtained from the SNO data alone. The bottom panel gives the allowed region including all existing data. The best fit point is:

\[
\Delta m_{12}^2 = 5 \times 10^{-5} \text{ eV}^2, \\
\tan^2 \theta_{12} = 0.32.
\]  

\[ (160) \]

9. REACTOR NEUTRINOS

Nuclear reactors are intense, isotropic sources of \( \bar{\nu}_e \) produced in their core as \( \beta \)-decay products of fission fragments (for a review and extensive references see [88]). In commercial nuclear reactors the energy is released in neutron induced fissions with a nuclear fuel constituted by uranium enriched (to 2–5 %) in the isotope \( ^{235}\text{U} \). The most important processes are of type

\[
n + ^{235}\text{U} \rightarrow X_1 + X_2 + 2n
\]  

\[ (161) \]

and result in the production of typically two neutrons (that can therefore sustain a chain reaction) and two fragments \( X_1 \) and \( X_2 \). The fragments of the fission are too rich in neutrons (a \( ^{235}\text{U} \) nucleus has 92 protons and 142 neutrons) and to reach stability must undergo a succession of beta decays (on average a total of six), therefore emitting an average of 6 electron anti-neutrinos. Considering that each decay releases 204 MeV, it is possible to obtain the number of anti-neutrinos emitted in a plant as:

\[
\tilde{N}_{\bar{\nu}_e} \approx 6 \times \left( \frac{\text{Power}}{204 \text{ MeV}} \right) \approx 8.9 \times 10^{20} \left( \frac{\text{Power}}{10 \text{ GigaWatt thermal}} \right) \bar{\nu}_e \text{ sec}^{-1}.
\]

\[ (162) \]

From a knowledge of nuclear physics it is possible to compute in detail the decay chains that are generated by the nuclear fissions, and estimate the flux and energy spectrum of the emitted \( \bar{\nu}_e \). The \( \bar{\nu}_e \)
energy is below 10 MeV, with an average value \( \sim 3 \) MeV. The neutrinos above the detection threshold
\( (E_{\nu_e} > 1.8 \) MeV\) are produced only in the decay of short lived fission fragments and therefore the flux
is very strongly correlated with the thermal power output of the nuclear reactor. For a detailed estimate
one has to consider the composition of the fuel, since four types of nuclei can undergo fission \( ^{235}U, ^{238}U, ^{239}Pu, \) and \( ^{241}Pu \). The chemical composition of the nuclear fuel slowly evolves with time, and this
must be also be taken into account. Taking into account all uncertainties a nuclear reactor \( \nu_e \) flux can be
estimated with an uncertainty of \( \pm 3\% \).

The \( \nu_e \) produced in the nuclear reactions can be detected with the reaction:

\[
\bar{\nu}_e + p \rightarrow e^+ + n.
\]  

(163)

This reaction has a threshold of 1.8 MeV and gives rise to two detectable signals: a prompt one from \( e^+ \)
production and the annihilation into two photons; and a delayed photon signal from neutron capture (after
neutron thermalization). The energy of the photon emitted in \( n \) capture is 2.2 MeV for the reaction \( np \rightarrow d\gamma \); it is often convenient to “dope” the scintillator with nuclei that have a large \( n \) capture cross section,
and result in higher energy photons, for example from \( n \) capture by Gadolinium one obtains \( E_\gamma \approx 8 \) MeV.

9.1. Chooz and Palo Verde

Recent important measurements of reactor \( \nu_e \) have been performed by two detectors: Chooz [89] (in
France), and Palo Verde [90] (in the USA) with a baseline of approximately 1 Km. A scheme of the
set-up of the Chooz detector is shown in Fig. 32. The source of neutrinos was composed of two nuclear
reactors, each with the thermal power of 4.2 GWatt. During a significant fraction of the data taking only
one of the reactors was in operation, and this allowed to study the backgrounds not-related to the neutrino
flux that are independent from the reactor power while the signal is linear with the instantaneous power
generated in the reactors cores.

The results of the Chooz experiments [89] are shown in Fig. 32 as the measured energy spectrum
of the detected positrons, compared to a calculated prediction in the absence of neutrino oscillations.
The energy spectrum extends from zero to 10 MeV with a maximum at approximately 3 MeV. The ratio

Fig. 31: Left panel: allowed areas from single solar neutrino experiments. The shaded area uses only Homestake data, while
the hatched area uses only the SNO charged-current rate. Overlaid (inside dashed lines) is the region allowed by Gallex/GNO
and SAGE. Right panel: allowed regions from a combined fit to these charged-current rates. All contours in this and other
figures are 95\% C.L.
between the measured spectrum and the theoretical prediction is:

\[
R_{\text{Chooz}} = \frac{\text{DATA}}{\text{Prediction}} = 1.01 \pm 2.8\% \text{ (stat)} \pm 2.7\% \text{ (syst)}.
\]  

(164)

The systematic error is due to uncertainties in the prediction of the neutrino fluxes (in the absence of oscillations)\(^{14}\). The Palo Verde experiment [90] has obtained a similar results:

\[
R_{\text{Palo Verde}} = 1.01 \pm 2.4\% \text{ (stat)} \pm 5.3\% \text{ (syst)}.
\]  

(165)

For both experiments the ratio Data/Prediction is consistent with unity. In the presence of oscillation the flux of \(\bar{\nu}_e\) at the detector would be suppressed by an energy dependent factor \(P_{\bar{\nu}_e \rightarrow \nu_e}(E, L)\), where \(L\) is the distance between the detector and the reactors nuclear core (the \(\bar{\nu}_\mu\) or \(\bar{\nu}_\tau\) generated by flavor transitions are below threshold to interact). The fact that no suppression of the flux is observed can be interpreted estimating an excluded region in the space of oscillation parameters. Qualitatively, it is clear that the result means that either the oscillation lengths for the transitions are longer than the distance \(L\) relevant for these experiments, or that the the oscillations can develop however their amplitude is sufficiently small that the effect remain unobservable. The simplest interpretation of the data is in terms of the “One mass scale dominance model”. This is equivalent to say that the smallest squared mass difference \((\Delta m^2_{12})\) is too small to have any influence, that is:

\[
\Delta m^2_{12} \lesssim \frac{(\hbar c)}{1 \text{ Km}} \approx 4 \times 10^{-4} \text{ eV}^2.
\]  

(166)

In this case the oscillation probability can be written as:

\[
P_{\bar{\nu}_e \rightarrow \nu_e} = 1 - \sin^2 2\theta_{13} \sin^2 \left[ 1.27 \frac{\Delta m^2(eV^2)}{E(\text{MeV})} \frac{L(\text{meters})}{E(\text{MeV})} \right].
\]  

(167)

This formula depends on two parameters: \(\Delta m^2 \simeq \Delta m^2_{23} \simeq \Delta m^2_{13}\) the ‘large’ neutrino squared mass difference and the mixing parameter \(\sin^2 2\theta_{13}\). As discussed in Section 4. one has that \(\sin^2 2\theta_{13} = |\langle \nu_e | \nu_3 \rangle|^2\), and therefore the oscillations vanish (that is \(\sin^2 2\theta_{13} = 0\) and \(P_{\bar{\nu}_e \rightarrow \nu_e} = 1\)) when \(\theta_{13} = 0\), that is when the state \(|\nu_3\rangle\) has no overlap with \(|\nu_e\rangle\), or when the \(\theta_{13} = \pi/2\) and the state \(|\nu_3\rangle\) coincides with \(|\nu_e\rangle\); however this last possibility is not consistent with the atmospheric neutrino data that give evidence for the

\(^{14}\text{Note how the detector has collected data until it has obtained a statistical error of the same order as the systematic uncertainty.}\)
Fig. 33: Neutrino Mixing with three parameters. (a) Allowed area of $\sin^2 \theta_{13}$ and $\sin^2 \theta_{23}$. At 90% C.L. SK data allows up to about 40% of $e$-type content of the atmospheric oscillation. (b) Allowed range of the atmospheric $\nu_e$ disappearance probability compared to the limits for $\nu_e$ obtained from various reactor experiments. The most stringent limit comes from the CHOOZ experiment.

The fact that the state $|\nu_3\rangle$ has large overlaps with $|\nu_\mu\rangle$ and $|\nu_\tau\rangle$. Equation (167) is formally identical to the two vacuum two flavor formula, and therefore a simple two-flavor analysis coincides with a three-flavor one performed the one mass scale approximation. Note also that equation (167) neglects the matter effects. In this case this is a good approximation, since for ordinary matter (with density $\rho \approx 2.7 \text{ g cm}^{-3}$) the effective potential is $V \approx 1.02 \times 10^{-13} \text{ eV}$, and for $E_\nu \lesssim 10 \text{ MeV}$, we have that:

$$2 E_\nu V \approx 10^{-8} \text{ eV}^2 \ll \Delta m^2$$

for all values of $\Delta m^2$ that can be explored. Since the experimental results are consistent with the absence of oscillations, the results can be interpreted as an exclusion region in the space of the parameters $\Delta m^2$ and $\sin^2 2 \theta_{13}$. These regions are shown in Fig. 33 together with the region that is allowed by the analysis of the atmospheric neutrino data in Super-Kamiokande [44]. The qualitative shape of the curves is simple to understand. Reactor experiments can exclude a region of high $\Delta m^2$ (corresponding to fast oscillations) and large $\sin^2 2 \theta_{13}$ (corresponding to large amplitude oscillations). For $\sin^2 2 \theta \approx 1$ $\Delta m^2$ must be smaller than a minimum value:

$$\Delta m^2_{\text{min}} \sim (hc) \left( \frac{3 \text{ MeV}}{1 \text{ Km}} \right) \approx 6 \times 10^{-4},$$

while for large $\Delta m^2$ the oscillation probability after the average over fast oscillations become a constant $P_{\nu_e \rightarrow \nu_e} \approx 1 - \frac{1}{2} \sin^2 2 \theta_{13}$, and one obtains a $\Delta m^2$ independent limit determined by the statistical and systematic error $\sin^2 2 \theta_{13} \leq 0.10$.

The atmospheric neutrino data determines an allowed interval for $\Delta m^2$ that, as discussed in Section 7., is determined by the observed distortion of the $\mu$-like data. Since the atmospheric neutrino data for $e$-like events are consistent with a no oscillation prediction, then the value $\sin^2 2 \theta_{13} = 0$ is allowed (and in fact corresponds to approximately the best fit point for the atmospheric neutrino data). The lack of measurable deviations from the no-oscillation prediction for atmospheric $e$-like events allows to set a limit on the maximum value of $\sin^2 2 \theta_{13} \lesssim 0.75$. As discussed in Section 7. this limit is not very stringent because of a cancellation between appearance and disappearance effects.

The combination of the results from atmospheric neutrinos (that determine the range of $\Delta m^2$) and reactor neutrinos allows to determine a lower limit on the value of $\sin^2 2 \theta_{13} \lesssim 0.25$; or equivalently

$$\sin^2 \theta_{13} \leq 0.065 \quad (90\% \text{ C.L.}).$$
9.2. The KamLAND detector

The evidence for the oscillations of solar neutrinos has been recently dramatically confirmed by the KamLAND experiment [91] with a measurement of the flux of reactor neutrinos for long ($L \sim 180$ Km) distance. The sensitive volume of the KamLAND detector (see Fig. 34) is a sphere of 6.5 meters radius that contains 1000 tons of liquid scintillator, instrumented with Photomultipliers. Electron anti-neutrinos can be detected with the usual reaction: $\nu_e + p \rightarrow e^+ + n$, measuring the positron in the liquid scintillator, the photon emitted in the neutron capture is also observable. The $\bar{\nu}_e$ flux is provided by an ensemble of approximately 70 reactors cores located at a distance between 150 and 210 Km from the detector in several commercial nuclear plants (see Fig. 34).

Recently the collaboration has published the results of the first 162 ton-year (145.1 days) of exposure. For this exposure the expected event rate was $86.8 \pm 5.6$ events (for the chosen threshold of 2.6 MeV). This has to be compared with a detected rate of 54 events, with a ratio

$$\frac{N_{\text{obs}} - N_{\text{BG}}}{N_{\text{expected}}} = 0.611 \pm 0.085(\text{stat}) \pm 0.041(\text{syst}).$$  (171)

The probability that this ratio consistent with unity is only 0.05%, and therefore the result is clear evidence for the existence of transitions $\bar{\nu}_e \rightarrow \bar{\nu}_e$. Fig. 35 shows the measured/expected ratio of different reactor neutrino experiments plotted as a function of the distance between the reactor and the detector. The solid dot is the KamLAND point plotted at the flux-weighted average distance (the dot size is indicative of the spread in reactor distances). One can clearly see that the ratio is compatible with unity (that is with the absence of oscillations) for pathlength $L \lesssim 1$ Km, while the KamLand ($L \sim 180$ Km) indicates that the survival probability $P(\nu_e \rightarrow \nu_e)$ is significantly different from unity. This result is in fact perfectly compatible with the Large Mixing Angle (LMA) solution of the solar neutrino problem. In Fig. 35 the shaded region indicates the range of flux predictions corresponding to the 95% C.L. LMA region found in the global analysis of the solar neutrino data. The dotted curve corresponds to $\sin^2 2\theta = 0.833$ and $\Delta m^2 = 5.5 \times 10^{-5}$ eV$^2$ that is a representative of recent best-fit LMA predictions.

Additional information on the possible existence of oscillations can be obtained from the analysis of the energy spectrum of the observed positrons. The expected spectrum is shown in the top part of the right panel of Fig. 35. One can see that the spectrum is dominated by the reactor neutrinos, with a smaller contributions of “geophysical” neutrinos (produced by radioactive decays of unstable nuclei in the Earth.
Fig. 35: The left panel shows the ratio measured/expected of the $\bar{\nu}_e$ flux from reactor experiments. The solid dot is the KamiLAND point. The shaded region indicates the range of flux predictions corresponding to the (95% C.L.) Large Mixing Angle Solution of the solar neutrino data. The dotted curve corresponds to $\sin^2 2\theta = 0.833$ and $\Delta m^2 = 5.5 \times 10^{-5}$ eV$^2$. The right panel shows on the top the expected reactor $\bar{\nu}_e$ energy spectrum with contributions of $\nu_{geo}$ and accidental background, and on the bottom the observed spectrum (solid circles with error bars), along with the expected no oscillation spectrum (upper histogram, with $\nu_{geo}$ and accidentals shown) and best fit (lower histogram) including neutrino oscillations. The shaded band indicates the systematic error in the best-fit spectrum. The vertical dashed line corresponds to the analysis threshold at 2.6 MeV.

The KamLAND results can be interpreted in terms of $\nu_e \rightarrow \nu_x$ transitions, obtaining an allowed region in the plane $(\sin^2 2\theta, \Delta m^2)$. The results for this allowed region are shown in Fig. 36. The figure shows in the usual plane $(\sin^2 2\theta, \Delta m^2)$ the region excluded by previous (shorter pathlength) reactor experiments Chooz and Palo Verde (all regions are for a 95% Confidence Level), and the allowed region of the LMA solution of the solar neutrino problem. The result of KamLand are shown in two ways: as an excused region, that takes into account only the total rate result (equation (171)); and an allowed region that takes into account both the total rate and the energy spectrum. This allowed region corresponds to several bands in $\Delta m^2$ that can reproduce the observed distortion of the spectrum. One can see that the KamLand data excludes all oscillations solutions for the solar neutrino problem with the exception of the LMA solution.

Several authors have performed combined analysis of the KamLand and solar neutrino data to determine the $\nu$ oscillation parameters. An example of this combined fits from Fogli et al. (92]) is shown in the right panel of Fig. 36. One can see that the LMA region is basically split into two sub-regions. The best fit point is in the lower region with $\Delta m^2_{12} \simeq 7.3 \times 10^{-5}$ eV$^2$ and $\sin^2 \theta_{12} \simeq 0.31$.  

180
Neutrino oscillations have been searched in accelerator experiments for a long time. Several dedicated searches for oscillation were motivated by cosmology and an analogy to the masses and mixing in the quark system. An explanation of the cosmological dark matter problem with neutrinos implies (see equation (36)) that the sum of the neutrino masses is of order $\sim 30 \text{ eV}$, on the other hand, an analogy with the quark sector suggests that $m_3 \gg m_2 \gg m_1$, and that the mixing angles are small. These considerations lead to the expectation that $\nu_{\mu, e} \leftrightarrow \nu_{\tau}$ oscillations could be detected with a large squared mass difference $\Delta m^2 \approx m_3^2 \sim (30 \text{ eV})^2$, but small mixing parameter $\sin^2 2\theta \ll 0.1$. Several dedicated experiment were performed searching for these oscillations. No evidence for these oscillations has been found. The best limits have been obtained by the NOMAD and CHORUS [94,95] detectors (see Fig. 37) using a neutrino beam produced by 400 GeV protons in the CERN laboratory. In these experiments the distance from the detector to the target region was of order 500 meters, the average neutrino energy was...
of order 50 GeV, with the minimum (threshold) energy for appearance \( E_\nu \simeq 3.5 \) GeV. The region of \( \Delta m^2 \) that can be explored is therefore:

\[
|\Delta m^2| \gtrsim (hc) \left( \frac{3.5 \text{ GeV}}{500 \text{ m}} \right) \simeq 0.7 \text{ eV}^2,
\]

while the very large statistics allows to study very small mixing \((\sin^2 2\theta \simeq 3 \times 10^{-4})\) at large \( |\Delta m^2| \).

The appearance of \( \nu_\tau \) charged current events can be detected using two techniques: (i) kinematical methods, or (ii) decay vertex identification methods. The NOMAD detector has searched for events with appropriate kinematical characteristic, in particular having missing transverse momentum, since in \( \nu_\tau \) charged current interactions some \( p_\perp \) is carried away by the neutrinos, that are always produced in \( \tau \) decay. The most attractive channel is the mode \( \tau^- \rightarrow \nu_\tau + e^- + \nu_e \), since it has a low background due to \( \nu_e \) charged-current interactions which are only \( \sim 1.5\% \) of the total event rate, and have a very different energy spectrum from what is expected from \( \nu_\mu \leftrightarrow \nu_\tau \) oscillations.

The CHORUS detector (see Fig. 38) was constructed using the emulsion technique that provides space resolution of \( \sim 1 \mu\text{m} \), well matched to the average \( \tau^- \) decay length of 1 mm.

Fig. 38: Expected configuration of a typical \( \nu_\tau \) charged-current interaction event in the CHORUS emulsion. The decay length for a \( \tau^- \) \((L_\tau \simeq 87 \mu\text{m} E_{\text{GeV}})\) is typically a fraction of a millimeter. In this example the \( \tau^- \) decays to \( \mu^- \nu_\tau \bar{\nu}_\mu \).

10.2. The LSND effect

A controversial result on neutrino flavor transitions has been obtained by the Liquid Scintillator Neutrino Detector (LSND) running at the Los Alamos Meson Physics facility \[96\]. The LSND collaboration has also reported evidence for neutrino flavor transitions of type \( \bar{\nu}_\mu \rightarrow \bar{\nu}_e \). This evidence is still very controversial, and has not been confirmed by other independent measurements. The “LSND effect” is not consistent with the results on solar and atmospheric neutrino experiments at least in a simple three neutrino framework, since it requires \( |\Delta m^2| \) in a range much larger than what is indicated by the other experiments. If this effect is real the consequences would very important.

The LSND \( \nu \) beam is obtained from an accelerated proton beam with momentum \(|\vec{p}| = 800 \) MeV, the detector is a tank filled with 167 tons of liquid scintillator at approximately 30 meters from the neutrino source. The neutrino beam is mostly due to \( \pi^+ \) that stop in target region and decay with the usual sequence \( \pi^+ \rightarrow \nu_\mu + \mu^+ \) followed by \( \mu^+ \rightarrow \bar{\nu}_\mu + \nu_e + e^- \). The contribution of \( \pi^- \) is much
suppressed because they are produced less abundantly in low energy $p$ interactions, and most of them are absorbed in nuclear interactions before decay, and therefore the beam contains $\nu_\mu$, $\bar{\nu}_\mu$, $\nu_e$ but only very few $\bar{\nu}_e$ (in the absence of oscillations the ratio $\bar{\nu}_e/\nu_e$ is estimated as $7.5 \times 10^{-4}$. The appearance of $\bar{\nu}_e$ can be studied with the “classic” coincident observation of a positron and a delayed photon from $n$ capture:

$$\bar{\nu}_e + p \rightarrow e^+ + n, \quad n + p \rightarrow d + \gamma (2.2 \text{ MeV}) .$$

(173)

After subtracting the expected sources of background The LSND collaboration [96] observes in the energy range $E_e \in [20, 60] \text{ MeV}$ an excess of $87.9 \pm 22.4 \pm 6.0$ events. This excess can be explained assuming an average probability:

$$\langle P_{\nu_\mu \rightarrow \nu_e} \rangle = (2.64 \pm 0.67 \pm 0.45) \times 10^{-3} .$$

(174)

This result is supported by a measurement of the $\nu_e$ flux with the process $\nu_e + ^{12}\text{C} \rightarrow e^- + X$, measuring electrons in the energy range $(60 \leq E_e \leq 200 \text{ MeV})$. Neutrinos in this energy range can only be produced by the decay in flight of pions and muons. For this class of events the LSND collaboration reports an excess of $18.1 \pm 6.6 \pm 4.0$ events consistent with $\nu_\mu \rightarrow \nu_e$ transitions with an average probability

$$\langle P_{\nu_\mu \rightarrow \nu_e} \rangle = (2.4 \pm 1.0 \pm 0.5) \times 10^{-3}$$

(175)

consistent with the previous result. Interpreting these results in terms of two flavor $\nu_\mu \leftrightarrow \nu_e$ oscillations one obtains the allowed region in the plane $(\sin^2 2\theta, \Delta m^2)$ shown in Fig. 39. The allowed region requires a mixing parameter $\sin^2 2\theta \simeq 2 \langle P_{\nu_\mu \rightarrow \nu_e} \rangle \simeq 4 \times 10^{-3}$ if $\Delta m^2$ is large. For maximal mixing ($\sin^2 2\theta \simeq 1$) the $\Delta m^2$ that can explain the data is of order

$$\Delta m^2 \gtrsim \frac{1}{3} (hc) \left( \frac{20 \text{ MeV}}{30 \text{ meters}} \right) \gtrsim 0.06 \text{ eV} .$$

(176)

The low $\Delta m^2$ region is however forbidden by the results of the Bugey reactor experiment. Note also that most of the allowed region of LSND is excluded by the results of the KARMEN [97], leaving only a small allowed interval.
Sterile Neutrinos?

The result of LSND cannot be easily reconciled with the results on atmospheric and solar neutrinos. This can be immediately understood noting that in three neutrino world there are only three independent squared mass differences $\Delta m^2_{jk}$, and the results

\[
\begin{align*}
\Delta m^2_{\odot} & \lesssim 10^{-4} \text{eV}^2, \\
\Delta m^2_{\text{atm}} & \simeq 3 \times 10^{-3} \text{eV}^2, \\
\Delta m^2_{\text{LSND}} & \gtrsim 0.2 \text{eV}^2
\end{align*}
\]

cannot be reconciled. The logical possibilities of course are that:

(a) one or more of the experimental results are incorrect;
(b) the theoretical framework used to interpret the results is incorrect or incomplete.

The addition of one or more neutrino mixed with the ordinary could explain the data, because in this case one more (independent) squared mass difference is available. The fourth light neutrino cannot be just the 4th generation neutrino similar to $\nu_e$, $\nu_\mu$ and $\nu_\tau$, because this would be in conflict with the experimentally measured width of $Z^0$ boson. It can only be an electroweak singlet (sterile) neutrino. The existence of such a state would of course be of the greatest importance.

Out of all experimental evidences for neutrino oscillations, the LSND result is the only one that has not yet been confirmed by other experiments. It is therefore very important to have it independently checked. This will be done by the MiniBooNE (first phase of BooNE) experiment at Fermilab. MiniBooNE will be capable of observing both $\nu_\mu \rightarrow \nu_e$ appearance and $\nu_\mu$ disappearance. The sensitivity of MiniBooNe is also shown in Fig. 39. MiniBooNE will begin taking data in 2002.

10.3. Long Baseline Experiments

In order to study neutrino oscillations with $\Delta m^2$ in the range suggested by the atmospheric neutrino data it is necessary to have baselines of order 100–1000 Km or more, depending on the neutrino energy. A comparison between a “near” and a “far” detector can be an ideal method to determined the existence and detailed properties of $\nu$ flavor oscillations. At present there is one project that is taking data: the K2K (or KEK to Kamioka) project, and two approved and under construction project: the Fermilab to Minos beam and the CERN to Gran Sasso beam.

In the K2K project an almost pure $\nu_\mu$ beam is generated with a 12 GeV proton beam in the KEK laboratory. The “far” detector is Super-Kamiokande at a distance of 250 Km.

The K2K experiment had a successful start in 1999 and has accumulated $\sim 4.8 \times 10^{19}$ “protons on target” (p.o.t.) that is approximately 40% of the goal of $10^{20}$ p.o.t. The time distribution of the contained events recorded during this period of data taking is shown in Fig. 40. The beam has the structure of a 1.1 $\mu$s pulse each 2.2 sec, and the analysis of fully contained events is virtually without background. Two “near” neutrino detectors are placed at the KEK site one being a one kiloton water detector that is like a miniature Super-Kamiokande, and the other a “fine-grained detector” made of scintillating fibers. This near detectors allow to monitor the intensity, stability and energy spectrum of the neutrino beam, when oscillations have not developed, and allow to extrapolate to the far (Super-Kamiokande detector).

The no-oscillation extrapolation of the near detector data predicts a rate of $80.1^{+6.2}_{-5.4}$ events in the fiducial volume of SK, while only 56 have been measured. The probability that the observed flux at SK is a statistical fluctuation is less than 1%.

The Fermilab to MINOS project will use the proton beam of the Main Injector at Fermilab
Fig. 40: Time distributions (relative to the beam arrival) of contained events in Super-Kamiokande during the data taking with the K2K $\nu$-beam. From [99].

Fig. 41: The left pane shows the total event count and single-ring $\mu$-like neutrino spectrum overlaid with expectations for no oscillations and oscillations at the K2K best fit point. The right panel shows the allowed region in the oscillation parameter space. From [99].
$E_\nu = 100$ GeV) to produce a neutrino beam that is sent to the underground site of MINOS at a
distance of 730 Km. The detailed comparison of a “near” and a “far” detector functionally identical
(two iron/scintillator sampling calorimeters) with toroidal magnetic field should allow to confirm the
oscillation interpretation for atmospheric neutrinos, and to determine more accurately the oscillation
parameters. The beginning of the data taking is scheduled for the end of 2004.

In the CERN to Gran Sasso project 450 GeV $p$ beam is the source of a higher energy neutrino beam $\langle E_\nu \rangle \sim 15 - 20$ GeV that will be sent to the Gran Sasso underground laboratory, again at a
distance of 730 Km. The OPERA detector is designed to search for the appearance of $\nu_\tau$ charged current
interactions with a massive lead/nuclear emulsion target. The ICARUS detector is also sensitive to the
$\nu_\tau$'s generated by the oscillations.

11. DOUBLE BETA DECAY

The most promising way to distinguish between Dirac and Majorana neutrinos is neutrinoless double
beta decay (for extensive reviews see [102]). Double beta decay is the process:

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- + 2\bar{\nu}_e \quad (2\nu\beta\beta \text{ decay}),$$

that can occur when single beta decay is kinematically forbidden. For example the nucleus $^{76}\text{Ge} (Z=32)$
cannot have a beta decay into the $Z=33$ state ($^{76}\text{As}$) that has a mass 0.4 MeV larger, but can have a double
beta decay into the $Z=34$ state ($^{76}\text{Se}$) that is 3.05 MeV lighter. The process (177) at the fundamental
(quark) level (see part (a) of Fig. 43) is the transition of two $d$ quarks into two $u$ quarks with the emission
of two electrons and two $\bar{\nu}_e$. The process is of second order in the weak coupling and therefore the
corresponding decay rates are very low with lifetimes of order $T > 10^{19} - 10^{21}$ years.

In the neutrino–less process:

$$(Z, A) \rightarrow (Z + 2, A) + 2e^- \quad (0\nu\beta\beta \text{ decay}),$$

there is no neutrino emission. The leading order diagram of this process is shown in part (b) of Fig. 43,
and can be pictured as one beta decay followed by the absorption of the emitted anti-neutrino by a
different neutron in the nucleus. The process has a very clear experimental signature because while in
the standard decay the sum of the energy of the two electrons in the final state has a broad distribution,
in the neutrinoless case one has that the sum of the energies of the two emitted electrons is equal to the
Fig. 43: Double beta decay diagrams. Diagram (a) shows the standard process when two neutrinos are emitted; diagram (b) shows the neutrinoless decay that has not been observed and is possible only if the neutrino is a massive Majorana particle.

$Q$ value of the reaction:

$$E(e_1^-) + E(e_2^-) \simeq Q \equiv M_i - M_f.$$  \hspace{1cm} (179) 

In order to have neutrinoless $2\beta$ decays one has two conditions:

1. The neutrino must be its own anti-particle, that is it must be a Majorana particle.
2. The neutrino mass must be non zero.

The first condition can be easily understood observing that the virtual neutrino in the diagram (b) of Fig. 43 is emitted as $\nu_e$ and is absorbed as $\bar{\nu}_e$, and in fact the reaction (178) clearly violates the lepton number by two units. The second condition follows from the fact that anti-neutrino emitted in the ‘first’ decay has the wrong chirality for being absorbed. The absorption is therefore possible without violating angular momentum conservation only if $m_\nu$ in non-vanishing. In fact the amplitude for neutrinoless double beta decay is proportional to the neutrino mass:

$$A(2\beta0\nu) \propto \sum_i U_{ei}^2 m_i \equiv \langle m_{\nu_e}\rangle_{\text{eff}}.$$  \hspace{1cm} (180) 

Note that the amplitude is proportional to a linear combination of the neutrino masses where the coefficients are $U_{ei}^2$ rather than $|U_{ei}|^2$, and therefore in general cancellations are possible, and the effective mass can be smaller than the smallest neutrino mass $m_j$. The quantity $\langle m_{\nu_e}\rangle_{\text{eff}}$ constitutes a lower limit on the mass of the heaviest neutrino.

The current best limits on neutrinoless double beta decay comes from Heidelberg–Moscow experiment on the double beta decay of $^{76}\text{Ge}$ [103]. This experiment located in the Gran Sasso underground laboratory is using 11 kg of Ge enriched to 86% in the isotope with mass 76, and has achieved a very low background (or order 0.2 counts/year/Kg/KeV) at the expected position of the $\beta\beta0\nu$ signal, and has collected an exposure of order 50 Kg years. In Fig. 44 we show the energy spectrum observed by the experiment (the experiment measures only the total released energy and cannot identify if the source of the energy is a photon, or one or more electrons). One can see the presence of several sources of background (that give a continuum and several radioactive decay lines) but also the clear contribution of a two-neutrino double decay with the predicted shape. The half-life of Germanium-76 was estimated from a fit to the background–subtracted signal in the energy range 700–2040 KeV

$$T^{2\nu}_{1/2} = [1.55 \pm 0.01\text{(stat)} \,^{+0.19}_{-0.15}\text{(syst)}] \times 10^{21}\text{years}.$$  \hspace{1cm} (181) 

In the presence of a neutrino–less contribution, one would observe a line corresponding to the $Q$-value of the $^{76}\text{Ge} \rightarrow ^{76}\text{Se}$ transition (2038.56 ± 0.32 KeV). The data (see the bottom panel of Fig. 44) shows
no peak above an approximately flat background. This can be used to extract a half-life limit for the neutrinoless mode

\[ T^{0\nu}_{1/2} \geq 1.9 \times 10^{25} \text{ years } 90\% \text{ C.L.} \] \hspace{1cm} (182)

This limits can be used to extract a limit\(^{15}\) effective mass of the \(\nu_\ell\) (see equation (180)):

\[ m_{\text{eff}} \leq 0.35 \text{ eV} \hspace{1cm} (90\% \text{C.L.}) \] \hspace{1cm} (183)

Recently it has been claimed \cite{104} that the results of the Heidelberg–Moscow collaboration can be interpreted as positive evidence for the existence of neutrinoless \(\beta\beta\) decay, and implies an effective neutrino mass \(\langle m_{\nu_\ell} \rangle_{\text{eff}} \approx 0.4 \text{ eV}\). This controversial claim has been criticised in \cite{105,106}.

The discovery of \(2\beta0\nu\) decay would establish that neutrinos are Majorana particle and determine the quantity \(\langle m_{\nu_\ell} \rangle_{\text{eff}}\) that sets a lower limit on the absolute value of the scale of the neutrino masses.

12. PHENOMENOLOGICAL SUMMARY

A coherent picture has begun to emerge from the experimental studies on neutrino flavor transitions. All existing data, with the notable exception of the LSND results, can be explained assuming the existence of standard oscillations among three neutrino flavors.

Without loss of generality one can define the state \(|\nu_3\rangle\) as the “most isolated neutrino”, that is the neutrino with the largest mass difference (in absolute value) from the closest mass eigenstates. Obviously \(|\nu_3\rangle\) is either the heaviest of the lightest neutrino. It is possible and convenient to choose \(\Delta m^2_{12} \equiv m_2^2 - m_1^2\) as positive (or \(m_2 > m_1\)). With this we have completely defined the conventions for the labeling of the neutrino mass eigenstates. The neutrino masses and mixing matrix have the following properties:

\[1\] The neutrino squared mass differences are ‘hierarchical’: there is a “small” mass difference \(\Delta m^2_{12}\), responsible for the oscillations of solar neutrinos, that is at least one order of magnitude smaller than “large” mass difference \(\Delta m^2_{23}\).

\(^{15}\)To obtain information of the effective mass parameter from the measured lifetime limit, one needs to use the results of a difficult calculation of the matrix element between the initial and final nucleus. This introduces some uncertainty in the limit
[2] The “large” squared mass difference is measures with atmospheric neutrinos and has value:

\[ \Delta m_{\text{atm}}^2 \simeq \Delta m_{23}^2 \simeq \Delta m_{13}^2 \simeq \pm 2.5 \times 10^{-3} \text{ eV}^2. \]  
(184)

Note the ambiguity in the sign. We do not know if \( \nu_3 \) is the lightest of heaviest neutrino. The 90\% C.L. allowed interval for \( |\Delta m_{23}^2| \) determined by SK is At 90\% C.L. the allowed interval is

\[ |\Delta m_{23}^2| \in [1.6, 4.5] \times 10^{-3} \text{ eV}^2. \]  
(185)

[3] The eigenstate \( |\nu_3\rangle \) has the composition:

\[ |\nu_3\rangle \simeq \frac{1}{\sqrt{2}}(1 \pm \epsilon) |\nu_\mu\rangle + \frac{1}{\sqrt{2}}(1 \mp \epsilon) |\nu_\tau\rangle + \epsilon' |\nu_e\rangle. \]  
(186)

where \( \epsilon \) and \( \epsilon' \) are small quantities. That is the “isolated” state \( |\nu_3\rangle \) is a superposition with approximately equal weight of \( \nu_\mu \) and \( \nu_\tau \) with a small (or vanishing) overlap with \( \nu_e \). In the standard parametrization of the mixing matrix we have:

\[ |\nu_3\rangle = \sin \theta_{13} |\nu_e\rangle + \cos \theta_{13} \sin \theta_{23} |\nu_\mu\rangle + \cos \theta_{13} \cos \theta_{23} |\nu_\tau\rangle, \]  
(187)

therefore we can restate the result (186) saying that \( \theta_{23} \simeq 45^\circ \) (that corresponds to maximal mixing) while \( \theta_{13} \) is small. For \( \theta_{23} \) the current 90\% C.L. allowed region determined by SK is \( \sin^2 2\theta_{23} \gtrsim 0.90 \) that can also be expressed as:

\[ \sin^2 \theta_{23}, \cos^2 \theta_{23} \in [0.35, 0.65] \quad (90\% \text{ C.L.}). \]  
(188)

[4] The upper limit on the value of \( \theta_{13} \) is determined by the results of the reactor experiments Chooz and Palo Verde and the non observations of \( \nu_e \rightarrow \nu_{\mu,\tau} \) transitions: The 90\% C.L. limits of Chooz are:

\[ \begin{align*}
\sin^2 \theta_{13} &\geq 0.065 \quad (\Delta m^2 = [\text{SK minimum allowed}] \}, \\
\sin^2 \theta_{13} &\geq 0.025 \quad (\Delta m^2 = [\text{SK best fit}] \}.
\end{align*} \]  
(189)

[5] The values of \( \Delta m_{12}^2 \) and of the other mixing angle \( \theta_{12} \) are determined by the solar neutrino experiments. The combination of the data of different experiments, until recently resulted in a set of disconnected regions in the plane \( (\Delta m_{12}^2, \sin^2 \theta_{12}) \), that have been given names: LMA corresponds to the “Large Mixing Angle” solution, SMA to the “Small Mixing Angle” solution, LOW to the region with large mixing angle with \( \Delta m^2 \) lower than for LMA, and finally the vacuum solution to the very small \( \Delta m^2 \) solution with oscillation length of the same order of the sun-earth distance. Numerically the regions are approximately:

\[
\begin{align*}
\Delta m_{12}^2 &= (0.7 \div 8) \times 10^{-5} \text{ eV}^2, \quad \sin^2 \theta_{12} = (0.2 \div 0.6) \quad \text{LMA} \\
\Delta m_{12}^2 &= (0.7 \div 2) \times 10^{-7} \text{ eV}^2, \quad \sin^2 \theta_{12} = (0.5 \div 0.7) \quad \text{LOW} \\
\Delta m_{12}^2 &= (0.4 \div 1) \times 10^{-5} \text{ eV}^2, \quad \sin^2 \theta_{12} = (0.6 \div 3.0) \times 10^{-3} \quad \text{SMA} \\
\Delta m_{12}^2 &= (0.5 \div 3) \times 10^{-10} \text{ eV}^2, \quad \sin^2 \theta_{12} = (0.25 \div 0.75) \quad \text{Vacuum}
\end{align*}
\]

The most recent data of SNO favors very clearly the LMA solution with a best point fit:

\[ \begin{align*}
\Delta m_{12}^2 &= 5 \times 10^{-5} \text{ eV}^2 \\
\sin^2 \theta_{12} &= 0.24 \quad (\theta_{12} = 29.5^\circ)
\end{align*} \]  
(190)

Note that since in the parametrization of the mixing matrix the overlap

\[ |\langle \nu_e | \nu_1 \rangle|^2 = \cos^2 \theta_{13} \cos^2 \theta_{12} \simeq \cos^2 \theta_{12}, \]  
the state \( |\nu_1\rangle \) state is mostly a \( \nu_e \).

These results indicate that the mixing in the lepton sector is deeply different than in the quark sector,
where all mixing angles are small. These results suggest (or better “require”) a number of new experimental studies to confirm and complete our knowledge of the neutrino masses and mixing. On the other hand the discovery of these remarkable “patterns” for the neutrino masses and mixing suggest several theoretical questions about what this all “means”, that is what it is teaching us about the origin of flavor, the fermion families and the particle masses.

Fig. 45: Flavor composition of the $\nu$ mass eigenstates. Each mass eigenstate is represented as a “box”, with three flavor components proportional to $|\langle \nu_e | \nu_j \rangle|^2$. The mixing angles are chosen as $\theta_{12} = 40^\circ$, $\theta_{23} = 45^\circ$, $\theta_{13} = 11^\circ$, the phase $\delta$ has the value $\delta = 90^\circ$. The value of $\theta_{23}$ is the best fit for the atmospheric neutrino data, $\theta_{12}$ is in the LMA solution for the solar neutrino data, while the value of $\theta_{13}$ is just below the upper limit obtained by the Chooz reactor experiment; the phase $\delta$ is left completely undetermined by the experimental data. The squared mass is only shown “qualitatively”. The ‘small’ mass splitting $m^2_{12}$ is responsible for (and determined by) the oscillations of solar neutrinos, while the larger splitting $m^2_{23} \approx m^2_{13}$ is responsible for the higher frequency oscillations measured with atmospheric neutrinos. Case (A) corresponds to “normal hierarchy”, case (B) to the “inverted hierarchy”.

12.1. Absolute value of the neutrino masses

Oscillation experiments can only determine squared mass differences, and therefore leave the question of the absolute value of the neutrino masses unanswered. Atmospheric ans solar neutrino experiments have determined the two independent $\Delta m^2$ with an important ambiguity in sign for $\Delta m^2_{atm}$. Using the notation $\Delta m^2_{\odot} \equiv \Delta m^2_{12}$ and $\Delta m^2_{atm} \equiv \Delta m^2_{23}$, and taking into account the sign ambiguity the neutrino masses can be written as:

- Normal hierarchy:

\[
\begin{align*}
    m_1 &= m_0, \\
    m_2 &= m_0 + \sqrt{\Delta m^2_{\odot}}, \\
    m_3 &= m_0 + \sqrt{\Delta m^2_{\odot} + \Delta m^2_{atm}}. 
\end{align*}
\]
Inverse hierarchy:

\[ m_3 = m_0, \]
\[ m_1 = m_0 + \sqrt{\Delta m_{\text{atm}}^2}, \]
\[ m_2 = m_0 + \sqrt{\Delta m_{\odot}^2 + \Delta m_{\text{atm}}^2}, \] (192)

where \( m_0 \) is an unknown minimal mass.

The experiments most sensitive to the mass \( m_0 \) are the searches for neutrinoless double beta decay. The amplitude for this process is proportional to an effective neutrino mass:

\[ m_{\text{eff}} = \left| \sum_j U_{ej}^2 m_j \right| \equiv \left| \sum_j |U_{ej}|^2 e^{-i\varphi_j} m_j \right|, \] (193)

where the phases \( \varphi_j \) are the Majorana phases that are unobservable in oscillation experiments. More explicitly, using the standard parametrization of the mixing matrix we have:

\[ m_{\text{eff}} = \left| c_{13}^2 c_{12}^2 m_1 + c_{13}^2 s_{12}^2 e^{i\varphi_2} m_2 + s_{13}^2 e^{i\varphi_3} m_3 \right|. \] (194)

13. MODELS FOR THE NEUTRINO MASS

The most striking fact about the neutrino masses is that they are so small, much smaller than the masses of the quarks and charged leptons. It is true that the fermion mass spectrum extends for several orders of magnitudes, however the size of the neutrino masses calls for a special explanation. A second remarkable fact is that the neutrino mass matrix is that the mixing is large, in sharp contrast with the quark sector where the Cabibbo–Kobayashi–Maskawa matrix \( V_{CKM} \) matrix has small deviations from the unity matrix.

Several explanations have been offered for the smallness of the neutrino masses. It is fascinating that all these explanations are based on some form of new physics beyond the standard model, and therefore the measurement of the neutrino masses and mixing can open a precious window on the physics of “Unification” giving information on the physics at mass scales beyond what can be reached directly with accelerator experiments.

A massive spin 1/2 particle has two spin states, and if it is “charged” it also has an anti-particle, with the same mass and opposite charge(s), that also has two spin states, for a total of four states with identical mass. Each quark or charged lepton (together with its own anti-particle) is in fact described as a 4-state system in a Dirac spinor. A “neutral” fermion on the other hand could be identical with its anti-particle, and have only two states. A CPT operation would transform the two states into each other. It is remarkable that it is possible to build the field theory of such a two state system described by a Majorana spinor.

Is the neutrino a Majorana or a Dirac particle? If the neutrino is massless this question is meaningless. This can be understood easily observing that a zero mass particle moves with the speed of light, and if it has helicity \(-1\) (that is spin anti-parallel to the momentum) in one system of reference it will have the same helicity in all frames, and of course the same is true for a +1 helicity particle. In the minimal standard model, each neutrino flavor has two states, the one with helicity \(-1\) is called by convention “neutrino” and the one with helicity +1 “antineutrino”. The question if the neutrino is a Dirac or Majorana particle is equivalent to the question if the reversal of the spin of a neutrino would turn it into an antineutrino or not. For a massless particle the spin reversal is impossible, and the question has no meaning. On the other hand if the neutrino has a non zero mass the question becomes not only meaningful but also natural and important. For a massive neutrino, with a Lorentz transformation, one
can “overtake” a neutrino produced in the laboratory with helicity $-1$, and study its behaviour in a new reference frame where it has the opposite helicity. If the boosted frame the neutrino behaves as a non-interacting singlet under the $SU(2) \otimes U(1)$ group it is the component of a four state Dirac spinor; if it behaves as an antineutrino it is the component of a two state Majorana spinor.

The possibility that the neutrino is a Majorana particle has suggested what is probably the most interesting explanation of the smallness of the neutrino mass, the “seesaw” mechanism. In the seesaw mechanism ([93]) (discussing for simplicity one single flavor) a four-state Dirac neutrino is split by “Majorana mass terms” into a pair of two-states Majorana neutrinos ($\nu$ and $N$). One of these Majorana neutrinos (the state $N$) becomes very massive taking a mass $m_N \sim M$ of the same order as a high mass scale $M$ characteristic of some new physics beyond the range of current experiments (for example a Grand Unification scale), while the other Majorana neutrino, that is identified with the observed light neutrinos, obtains a light mass $m_\nu$ that is of order $m_\nu \sim m_D^2 / M$ where $m_D$ is the Dirac mass characteristic of the other fermions in the same generation. So the reason why the neutrino is light is that it is associated with a very massive particle, in the same way as in the game of the seesaw (see Fig. 46) a massive participant will push “up” the other one.

![Fig. 46: Scheme of a see-saw.](image)

A technical and detailed description of this mechanism is beyond the scope of these lectures (see [10, 4] for more discussion), however the basic formulae in the model are very simple and instructive. A 4-components Dirac spinor can be represented as the combination of a pair of two-components spinors (the left-handed and right-handed part). In the standard (Weyl of chiral representation) one has:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$$

where $\psi_{L,R}$ are the two components Weyl spinors with left and right chirality. The operation of particle–antiparticle conjugation for a Dirac spinor corresponds to $\psi \rightarrow \psi^c$ with:

$$\psi^c = C \overline{\psi}^T = i \gamma^0 \gamma^2 \psi^T = i \gamma^2 \psi^* = \begin{pmatrix} +i\sigma^2 \psi^*_R \\ -i\sigma^2 \psi^*_L \end{pmatrix}$$

where $\overline{\psi} = \psi^\dagger \gamma^0$, the asterisk indicates complex conjugation, and we have used the explicit form of the Dirac matrices in the chiral representation and $\sigma^2$ is the second one of the three Pauli matrices. Note that the conjugate of a left-handed spinor has only a right-handed component and viceversa, in fact particle–antiparticle conjugation transform a left (right) handed particle in a right (left) handed anti-particle; it is also easy to verify that $(\psi^c)^c = \psi$. The standard Dirac mass term in the Lagrangian has the form:

$$-L_{\text{mass}}^{\text{Dirac}} = m_D \overline{\psi} \psi = m_D \left[ \psi_L^\dagger \psi_R + \psi_R^\dagger \psi_L \right]$$

and combines the left and right handed parts of the spinor. A Majorana particle with only two degrees of freedom is described by one single Weyl spinor, and therefore a Dirac mass term such as (197) cannot
exist, however one can still construct a different mass term. If $\psi_L$ is a left-handed spinor, one can obtain an object that transform as a right handed one as: $(\psi_L)^c = -i\sigma^2 \psi^*_L$. Note that one can then represent a Majorana spinor also as a 4-component object as:

$$\psi_M = \begin{pmatrix} \psi_L \\ -i\sigma^2 \psi^*_L \end{pmatrix} = \begin{pmatrix} \psi_L \\ (\psi_L)^c \end{pmatrix}.$$  \hspace{1cm} (198)

This form has obviously only two independent components and under a charge conjugation operation (as described in equation (196)) one has $(\psi_M)^c = \psi_M$. A mass term with a single Weyl spinor (left or right-handed) can then be constructed as:

$$-\mathcal{L}_{\text{mass}}^{\text{Majorana}} = \frac{m_M}{2} \{ (\psi_L)^c \dagger \psi_L + \text{h.c.} \} = \frac{m_M}{2} \{ (\psi_L)^T i\sigma^2 \psi_L + \text{h.c.} \}. \hspace{1cm} (199)$$

Such a Majorana mass term violates (by two units) the conservation of the “charges” associated to the fermion considered, therefore it is only possible for “neutral” particles. Since neutrinos have no known charges, in principle they can have a Dirac mass term \textit{and} a Majorana mass term. One can write this combination in matrix form as:

$$-\mathcal{L}_{\text{mass}}^\nu = \frac{1}{2} [\nu_L \ (\nu_R)^c] C \begin{pmatrix} m_{M,L} & m_D \\ m_D & m_{M,R} \end{pmatrix} \begin{pmatrix} \nu_L \\ (\nu_R)^c \end{pmatrix} + \text{h.c.}, \hspace{1cm} (200)$$

where we have included a Dirac mass term, and two Majorana masses for the left and right component. Now it is possible to observe that:

1. The Dirac mass term can be naturally expected to be of the same order of magnitude as for the other fermions in the family.
2. The mass term $m_{M,L}$ is small ($m_{M,L} \ll m_D$), because it can be generated only by a triplet of Higgs scalars, that is absent in the standard model.
3. A right-handed neutrino is completely neutral under the standard–model gauge group, and therefore it can have any mass at all without breaking the gauge symmetry. In other words the mass $m_{M,R}$ is not connected to the symmetry breaking scale ($v = 246$ GeV) of $SU(2) \otimes U(1)$ but is associated to a different higher mass scale $M$ (connected for example with unification).

We therefore have a mass matrix of type:

$$\mathcal{M} \simeq \begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix}. \hspace{1cm} (201)$$

This matrix can be diagonalized, obtaining two eigenvalues that are approximately $M$ and $-m_D^2/M$ (the minus sign is irrelevant and can be reabsorbed by a redefinition of the neutrino field). The two eigenvectors are two physical Majorana Particles with masses $m_N \simeq M$ and $m_\nu \simeq m_D^2/M$. In summary we have used two ingredients: (i) the possibility to include a Majorana like mass term, and (ii) the existence of a large mass scale $M$, to obtain the result that the neutrino splits naturally into two Majorana states one very massive and one very light. As an illustration, one could extract a new mass scale from the atmospheric neutrino data using:

$$|\Delta m^2_{\text{atm}}| \simeq m_2^2 - m_1^2 \simeq m_3^2 \simeq \left( \frac{m_{\text{top}}^2}{M} \right)^2, \hspace{1cm} (202)$$
obtaining:

\[ M \sim \frac{m_{\text{top}}^2}{\sqrt{\Delta m_{\text{atm}}^2}} \sim 0.6 \times 10^{15} \text{ GeV}, \]  

(203)

that is close to the Grand Unification scale\(^\text{16}\).

More in general, considering the case with \( n \) generations of neutrinos, the mass matrix (201) becomes of order \((2n \times 2n)\) (each entry has 2 components). The diagonalization of this matrix results in \( n \) light Majorana neutrinos and \( n \) super heavy ones, however now the light neutrinos are mixed with each other (with negligible admixtures with the heavy ones of order \( m_D/M \)). The mixing clearly depends on various assumptions about the masses. Can this procedure reproduce the patterns of mixing observed in nature? The original predictions of lepton mixing in the framework of the seesaw model predicted that in analogy with the quark sector, the mixing between light neutrinos would be small, and the experimental results came as big surprise for (essentially all) the theory community. It is now understood that the seesaw model can predict also large mixing. Several non-seesaw models for the neutrino masses also exist, also making use of physics beyond the standard model. This is a very active field of research that cannot be summarized here (see for example [93] for a discussion and additional references). We list below some possible questions for model builders:

1. Why some of the \( \nu \) mixing angles are large while all quark mixing angles small?
2. Why two of the \( \nu \) angles are (probably) large while one (\( \theta_{13} \) is small)?
3. Why \( \theta_{23} \) is so close to the “maximal” value of 45\(^\circ\)? How close it is to this special value?
4. How large is \( \theta_{13} \) ?
5. Are the neutrino masses nearly degenerate with small splittings or “hierarchical”?
6. Is the ordering of the neutrino masses “normal” (with the state \( \nu_1 \) that is prevalently \( e \)-like as the lightest state) or “inverted”?
7. Is the phase \( \delta \) predictable?

Research on these (and other) questions is very active, and hopefully the answers will shed light on the physics of unification.

14. CONCLUSIONS

The field of neutrino Physics has lived during the last decade a period of remarkable developments, and the perspectives for future studies are exciting. Future solar and long-baseline experiments will provide new and important knowledge, and should determine the neutrino squared mass differences and mixing; double beta decay studies and/or cosmological studies can also determine the absolute scale of the \( \nu \) masses. Neutrinos may also bring us new surprises, as they have done several times in the past, revealing some new properties, or the existence of additional states. Theoretically, the challenge of understanding the deeper meaning of the neutrino properties in a unification scheme remains an open problem of fundamental importance.

Acknowledgments

It was a great pleasure to participate to the CLAF school, and I’m very grateful to the students and the organizers for the stimulating and pleasant atmosphere. A special thanks to prof. Bernard Marechal, Maurizio Lusignoli, Carlo Bernardini and Andrea Donini for criticism and discussions.

\(^{16}\)Replacing \( m_{\text{top}} \) with \( m_T \) reduces the scale \( M \) by 4 orders of magnitude to \( M \simeq 6 \times 10^{11} \text{ GeV} \), and one can speculate about the significance of this new mass scale.
References


[27] http://www.sdss.org/


[37] C. Giunti and C.W. Kim, hep-ph/0011074; see also http://www.to.infn.it/~giunti/neutrino/qmno/


[86] SNO collaboration, nucl-ex/0204008.

[87] SNO collaboration, nucl-ex/0204009.


[100] See the MINOS home page: http://www-numi.fnal.gov/

[101] See the CNGS home page: http://proj-cngs.web.cern.ch/proj-cngs/


Abstract
The field of ultra-relativistic heavy ion physics, which started some 15 years ago at the Brookhaven AGS and the CERN SPS with fixed target experiments, is entering today a new era with the recent start-up of the Relativistic Heavy Ion Collider RHIC and preparations well under way for a new large heavy ion experiment at the Large Hadron Collider LHC. This overview will sketch a rough picture of the heavy ion program at current and future machines and concentrate on a few important topics, in particular the question if current results show any of the signs predicted for the phase transition between normal hadronic matter and the Quark-Gluon Plasma.

1. INTRODUCTION

The aim of high-energy heavy-ion physics is the study of strongly interacting matter at extreme energy densities. Statistical QCD predicts that, at sufficiently high density, there will be a transition from hadronic matter to a plasma of deconfined quarks and gluons – a transition which in the early universe took place in the inverse direction some $10^{-5}$ s after the Big Bang and which might play a role still today in the core of collapsing neutron stars. The study of the phase diagram of nuclear matter (see Fig. 1), utilising methods and concepts from both nuclear and high-energy physics, constitutes a new and inter-disciplinary approach in investigating matter and its interactions. In high-energy physics, interactions are derived from first principles (gauge theories), and the matter concerned consists mostly of single particles (hadrons/quarks). In contrast, on nuclear physics scales the strong interaction is shielded and can, therefore, to date only be described in effective theories, whereas matter consists of extended systems with collective features. Combining the elementary-interaction aspect of high-energy physics with the...
macroscopic-matter aspect of nuclear physics, the subject of heavy-ion collisions is the study of bulk matter consisting of strongly interacting particles (hadrons/partons), i.e. QCD thermodynamics.

The study of the QGP is of interest to explore and test QCD on its natural scale ($\Lambda_{QCD}$) and addresses the fundamental questions of confinement and chiral-symmetry breaking, which are connected to the existence and properties of the quark-gluon plasma. Moreover, it is of general relevance in understanding the dynamical nature of phase transitions involving elementary quantum fields, as the QCD phase transition is the only one accessible to laboratory experiments.

2. CURRENT STATUS AND RESULTS

The study of ultra-relativistic heavy-ion collisions is a rather new, but rapidly evolving field. After the pioneering experiments at the BEVALAC and in DUBNA with relativistic heavy ions ($E/m \approx 1$), the first experiments started in 1986 with light ions almost simultaneously in Brookhaven (AGS) and at CERN (SPS). Really heavy ions ($A \approx 200$) have been available in the AGS since the end of 1992 and at the SPS since the end of 1994.

The year 2000 was an extremely important and fruitful one for ultra-relativistic heavy ion physics. It started of with an appraisal of the CERN SPS Pb beam results [1] which concluded that 'compelling evidence has been found for a new state of matter' featuring many of the characteristics expected for a Quark–Gluon Plasma (QGP). Later that year the RHIC collider and its four experiments at BNL began operation with Au–Au collisions at up to 130 GeV/nucleon. Already this first short run was a success beyond even the most optimistic expectations, with physics results appearing in print and essentially flooding the relevant conferences within months. It is therefore timely to both assess the results achieved at the fixed target machines and to ask what, if anything, the LHC heavy ion program can contribute beyond existing facilities to the study of strongly interacting matter under extreme conditions.

2.1. Initial conditions and global features

The predictions of lattice QCD are rather firm in that a transition to the QGP should exist in the vicinity of a critical temperature $T_c$ of $150 – 200$ MeV (whether the transition is of first order, second order, or only 'rapid' is still a matter of debate). However, whether the QGP is actually created in heavy-ion collisions at current energies is a different question and will depend on the dynamics of the reactions and in particular on the initial conditions of the system shortly after the collision. In order to reach the QGP, or even only to use macroscopic concepts (such as 'phase transition') and the language and variables of thermodynamics (such as 'temperature' or 'density'), the system has to be extended – i.e. its dimensions ought to be much larger than the typical scale of strong interactions — it has to be in (or near) equilibrium – i.e. its lifetime has to be larger than the relevant relaxation times – and the energy density $\epsilon$ has to exceed the critical threshold for QGP formation. This threshold is predicted by lattice QCD to be of the order of $1 – 3$ GeV/fm$^3$, equivalent to a temperature $T_c$ of $150 – 200$ MeV or a baryon density $\rho_c$, of 5 to 10 times normal nuclear matter density (see Fig. 1). Present results from the ongoing fixed-target program indicate that the initial conditions realized in these reactions could indeed be favourable for QGP formation. In head-on central collisions, hundreds of particles are produced per unit of rapidity, the system expands to a size of the order of 1000 fm$^3$ (as measured by particle interferometry), and initial energy densities are estimated to exceed 2 GeV/fm$^3$. However, the expansion is also extremely fast, with an estimated total lifetime of only a few fm/c from the first instance of the collision until the final freeze-out of hadrons.

While these results show that we are certainly close to the requirements listed above for QGP formation, they are by no means sufficient. In particular the energy density estimates are inversely proportional to the assumed 'formation time', i.e. the time needed to reach thermal equilibrium, and might
well be smaller (or bigger?) by a factor of the order of two. Also, the lifetime of the system seems marginal, and even if a QGP is formed it might simply not live long enough for its signals to clearly stand out from the background created in later, hadronic phases of the evolution. The existence of a QGP phase can only be settled experimentally by searching for direct and specific signals.

2.2. Recent experimental highlights

The following sections will concentrate on three main topics which are at the heart of the quest for the QGP, and in which significant progress has been achieved over the last years, i.e. are there experimental indications for *equilibrated hadronic matter*, *chiral symmetry restoration*, and *deconfinement*?

**Equilibrium hadronic matter?** While in principle the study of non-equilibrium hadronic matter might be of considerable interest, in practice the huge number of largely unknown dynamical parameters governing the evolution of heavy ion reactions would make the analysis of such a complex system very difficult. The powerful laws of thermodynamics can reduce this complexity and make definite and testable predictions, largely independent of the microscopic dynamics, for those degrees of freedom which evolve in equilibrium. The price to pay is a loss of information concerning events preceding the equilibrium, as the memory of earlier (and possibly more interesting) stages of the evolution is largely lost.

In reality, like during the evolution of the early universe, we will have to deal with a hierarchy of processes and scales. Some of these have large cross-sections and correspondingly small relaxation times and therefore might evolve close to equilibrium, and others will decouple early from a thermal evolution and are sensitive to the hot initial phase of the reaction. Prime candidates for the former are hadronic observables, like momentum spectra and particle ratios, and for the latter hard probes and electromagnetic signals.

In a purely thermal system of hadrons, the momentum distributions, when expressed as a function of the transverse mass $m_T (m_T = \sqrt{m^2 + p_T^2})$, will be independent of the particle mass with a slope inversely proportional to the temperature $T$. In an expanding system, an additional collective flow component can develop which blue-shifts the momentum spectra with a common transverse velocity $\beta_T$ leading to a mass dependent component. Likewise, the abundance of particle species in equilibrium hadronic matter is given by two independent parameters, i.e. the temperature $T$ and a baryochemical potential $\mu_B$ (which reflects the baryon asymmetry in the initial state). A hadronic system in both ‘thermal’ (momentum) and ‘chemical’ (particle abundance) equilibrium is therefore fully determined by only three independent parameters: $T$, $\beta_T$ and $\mu_B$.

Such a simple prescription seems to be indeed borne out by the data. This is illustrated in Fig. 2, which shows a comparison of measured particle ratios with predictions based on chemical equilibrium [2]. These ratios are in rather good agreement with the equilibrium calculations for a temperature of about 170 MeV and a baryon chemical potential of $\approx 250$ MeV, corresponding to 1/3 nuclear matter density (at the AGS the corresponding values are $T \approx 140$ MeV and $\mu_b \approx 500$ MeV).

Momentum spectra of different particles in Pb-Pb reactions [3] are also well described by a thermal distribution, if, in addition, a common (to all particle types) flow velocity of $\beta_T \approx 0.4 \, c - 0.6 \, c$ is introduced (see Fig. 3). On the right part of Fig. 3, the inverse $m_T$-slopes are shown for pp, S+S and Pb+Pb reactions at comparable energies ($\sqrt{s} \approx 20$ GeV). While the slopes in pp reactions are independent of particle type, i.e. exhibit ‘$m_T$-scaling’, the slope parameter increases proportional to the particle mass for heavier reaction systems as expected for collective flow. The temperature extracted from momentum spectra at SPS is of the order of 120 MeV, i.e. significantly lower than the one extracted from the particle ratios mentioned above.
Fig. 2: Hadrochemical equilibrium model calculation of hadron yields (full lines, calculated for a temperature $T$ of 168 MeV and a baryochemical potential $\mu_B$ of 266 MeV) compared to data from CERN SPS for Pb+Pb at 158 GeV/c.

Fig. 3: Transverse mass spectra (left) and slope parameters (right) of pions, kaons and protons near midrapidity from NA44.

A large set of independent hadronic observables, i.e. momentum spectra, particle ratios and HBT correlation results (which are also sensitive to $T$ and $\beta_T$ [4]), seems to be consistent with a surprisingly simple picture of the late stages of heavy ion reactions: different particle species are created in relative abundance consistent with chemical equilibrium ratios at a 'temperature' of $T \approx 170$ MeV; this dense hadronic system then expands and cools to a temperature of about 120 MeV, converting random 'thermal' motion into ordered collective flow until the final freeze-out, when the system is so dilute that all interactions cease. This experimental phase diagram, with both chemical and thermal freeze-out locations as determined from data ranging from very low energies (SIS) up to SPS, is shown in Fig. 4 [5]. The location of the particle ratio freeze-out point in the temperature-density plane is located very close to the expected phase boundary between hadronic matter and the QGP, and the distance between chemical and thermal freeze-out increases with the beam energy, indicating an increasing dynamical path in the hadronic phase for larger systems (more final state particles).

However, before this intuitively appealing scenario can be taken as fully established, a number of experimental and conceptual questions will have to be clarified. On the experimental side, resolving some inconsistencies between different experiments and better statistics over a large range of impact parameters (in particular for Hyperons) will be needed to come to a more quantitative test of predic-
Fig. 4: Compilation of chemical (particle ratios) and thermal (momentum spectra) freeze-out points from SIS to SPS energies. On the conceptual side, the most puzzling observation is that already very elementary reactions look practically as 'thermal' as heavy ion reactions. While it has been known, but never 'understood', that momentum spectra in hadronic reactions look 'thermal' (obey $m_T$-scaling, see Fig. 3), a recent re-analysis of pp and $e^+e^-$ reactions has shown that also the particle ratios can be extremely well described with thermodynamics [6]. How can this be possible in systems containing only a few hadrons where a dynamical path from arbitrary initial conditions to thermal distributions via interactions (rescattering) is very unlikely? The success of thermal models to describe particle ratios in reactions ranging from $e^+e^-$ at LEP to Pb-Pb at the SPS could be a hint for a universal feature of the parton-to-hadron (phase?) transition, which might be governed by statistics and phase space at the time of particle creation rather than by dynamical features [5,7].

Assuming that some satisfactory answers to these questions can eventually be found, we could then go on to analyse the hadronic data in more detail to look for information on the dynamics preceding the freeze-out. Relaxation times in a partonic and a hadronic medium are likely to be different, and therefore the questions how and how fast did the system reach equilibrium in different channels are of interest, particularly in the Hyperon sector, where hadronic relaxation times are estimated to be extremely long. Flow patterns should be sensitive to the equation of state of matter and therefore contain indirect evidence for a QGP phase transition preceding freeze-out.

**Chiral symmetry restoration?** Weakly interacting electromagnetic probes (photons or leptons) are a direct means of gaining information on the early dense and hot stages of the collision, as they leave the interaction volume without being altered by final state effects. While, so far, only upper limits exist for direct (thermal) photon production, recent data on lepton pairs show an unexpectedly large yield at low masses, below the $\pi^0$ meson.

Figure 5 shows the electron pair mass spectrum observed in central S+Au collisions by NA45 [8]. The upper part summarizes model calculations which include contributions from hadronic decays (shaded area) and from in-medium pion annihilation and bremsstrahlung. An excess at 0.2 <
$m(e^+e^-) < 0.6$ remains unexplained. The lower panel exhibits perfect agreement with the data obtained in models which include in addition an in-medium modification of the $\varphi$ and $\omega$ masses. A similar excess, consistent with the same model calculations, has been found in the $\mu^+\mu^-$ mass spectrum by NA34/3 [8].

![Graph](image1)

**Fig. 5:** Di-electron invariant mass distribution measured by the NA45 experiment in central S+Au collisions, compared to calculations including hadronic decays and effects expected for high pion densities (top) and calculations incorporating in addition a density dependent mass shift of the $\varphi$ meson (bottom).

![Graph](image2)

**Fig. 6:** $J/\Psi$ production for proton, sulphur and lead induced collisions relative to the Drell-Yan yield as a function of the thickness L of matter traversed on average. The data are divided by the normal nuclear absorption (exponential in L) which is consistent with the results up to peripheral Pb+Pb reactions. A sudden onset of an additional 'anomalous' suppression is observed for central Pb+Pb.

In-medium modification of vector mesons, if experimentally confirmed by more conclusive data, could be a direct consequence of the chiral symmetry transition at the phase boundary between hadronic matter and the QGP. The rapidly varying quark condensate should lead to changes in the properties of hadrons (masses, width) in the vicinity of the phase transition, which will be observable in the lepton mass spectrum for mesons decaying in the dense transition regime. This would indeed be a spectacular verification of the concept underlying the generation of light hadron masses in QCD.

An excess in the intermediate mass range (1.5 – 2.5 GeV) observed in muon pairs by NA50 has so far not found any convincing interpretation [9]. Speculations concerning its origin range from enhanced open charm production and final state rescattering of produced charm quarks to thermal radiation of virtual photons. The new experiment NA60 is currently under preparation at the SPS to address this question.
Deconfinement? Signals originating from hard-scattering processes at the very beginning of the reaction are an ideal tool to probe the state of the surrounding QCD matter. The original idea [10] that $J/\Psi$ production should be suppressed in a QGP relies on a Debye screening mechanism which renders colour interactions short ranged in a dense medium (‘deconfinement’) and therefore prevents the formation of bound resonances.

$J/\Psi$ suppression with similar characteristics as predicted for a QGP was indeed one of the first results reported from heavy ion experiments in 1987. Its subsequent interpretation, alternating repeatedly between ‘trivial’ and ‘exciting’, is probably the best example on how our understanding of nuclear collisions has progressed in a constant interplay between theory and experiment, new explanations and new data. A compilation of $J/\Psi$ production relative to the Drell-Yan continuum in pA and AB reactions is shown in Fig. 6 versus the average path length $L$ traversed by the $c\bar{c}$ pair after its creation inside the target and projectile nuclei [11]. Up to and including central S-U collisions, this ratio decreases exponentially with $L$, consistent with a nuclear absorption cross-section of 6 mb. It took the better part of the last ten years, a variety of data for $J/\Psi$, $\Psi'$ and $\Upsilon$ – from low energy pp and pA reactions to photoproduction and high $p_T$ production at the Tevatron – and a good measure of other ingredients (nuclear structure functions, initial and final state scattering, formation time) to come to a consistent and theoretically substantiated interpretation [12,13]. The exponential attenuation is today seen as resulting from the interaction between the nuclear medium and a pre-resonance state, a coloured $c\bar{c}$-gluon configuration which evolves only later (and outside the nucleus) after some finite formation time into the physical, colour neutral $J/\Psi$ or $\Psi'$ hadron. So $J/\Psi$ suppression has provided a lot of insight into the dynamics of charmonium production, hadron formation and using the nucleus as a tool to measure short time scales, but leaving no room for QGP effects.

The extrapolation of this model to central Pb-Pb reactions was straightforward, essentially parameter free, and completely wrong (see Fig. 6)! The Pb-Pb data, whether plotted as a function of $L$ or any other variable, shows significantly less $J/\Psi$’s than hadronic absorption models would predict by extrapolating from light ion and pA results. While some debate still persists if the additional suppression is really ‘anomalous’ or not, new precision data which is currently being analysed should settle this question in the near future. Most likely, some additional physics will have to be included in order to describe the Pb data. However, whether this ‘new physics’ will require deconfinement, dynamical pre-cursor phenomena of the QGP transition, or just some overlooked hadronic effect, remains to be seen.

2.3. Future fixed target program

Given the recent exciting developments, the future directions are perfectly clear. The SPS fixed target program has entered an extremely productive phase and it is now being brought to its full potential. With the exception of the Hyperon and the low mass lepton pair measurements, statistics is, in general, not a problem. A run at the lowest possible SPS energy, around 40 GeV/nucleon and still being analyzed, has increased the baryon density, possibly close to its maximum value. Signals related to chiral symmetry restoration, in particular the low mass lepton pairs, will in general be rather sensitive to baryon density. The low energy run can also make contact with the AGS regime and will allow the CERN experiments, which are quite distinct in their capabilities from the AGS detectors, to compare with and complement the program at lower energies. The SPS program has been extended beyond the year 2001 to address in particular charm production and the intermediate mass lepton pair excess.
3. HEAVY ION PHYSICS OF THE 21ST CENTURY

With the colliders RHIC ($\sqrt{s} = 200$ GeV/n) and LHC ($\sqrt{s} = 5.5$ TeV/n) coming into operation in 2000 and 2006, respectively, the available energy in the centre-of-mass will have increased by almost five orders of magnitude within 20 years. This unprecedented pace was made possible only by (re)using accelerators, and to some extent even detectors, built over a much longer time scale for use in high-energy physics. The following sections will summarise the physics and experiments to come in these latest (and possible last) heavy-ion machines.

3.1. Initial conditions and global event features

In order to get some qualitative feeling about the changes in global event features to be expected when going up in energy by over two orders of magnitude from SPS via RHIC to LHC, some of the relevant parameters and predictions are listed in Table 1 and described below:

Table 1: Qualitative extrapolation for particle production from pp to AA and estimated global features in AA. For details see text.

<table>
<thead>
<tr>
<th></th>
<th>SPS</th>
<th>RHIC</th>
<th>LHC</th>
<th>Pb–Pb</th>
<th>SPS</th>
<th>RHIC</th>
<th>LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{s}/A$ (GeV)</td>
<td>17</td>
<td>200</td>
<td>5500</td>
<td></td>
<td>400</td>
<td>800</td>
<td>2-8x10^3</td>
</tr>
<tr>
<td>pp $dN_{ch}/dy$</td>
<td>1.8</td>
<td>2.4</td>
<td>5</td>
<td>$\tau_0^{QGP}$ (fm/c)</td>
<td>1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>pp $p_t$ (GeV)</td>
<td>0.36</td>
<td>0.39</td>
<td>0.53</td>
<td>$\varepsilon$ (GeV/fm$^3$)</td>
<td>3</td>
<td>35</td>
<td>500</td>
</tr>
<tr>
<td>AA $dN/dy$</td>
<td>$\propto N_{part}$</td>
<td>$\propto N_{col}$</td>
<td>$\tau_{QGP}$ (fm/c)</td>
<td>$\lesssim 2$</td>
<td>2-4</td>
<td>$\gtrsim 10$</td>
<td></td>
</tr>
<tr>
<td>partons in p</td>
<td>4</td>
<td>10</td>
<td>30</td>
<td>$\tau_{fo}$ (fm/c)</td>
<td>10</td>
<td>20-30</td>
<td>30-40</td>
</tr>
<tr>
<td>shadowing</td>
<td>1</td>
<td>0.8</td>
<td>0.5</td>
<td>$V_{loc}$ (fm$^3$)</td>
<td>few 10$^3$</td>
<td>few 10$^4$</td>
<td>few 10$^5$</td>
</tr>
<tr>
<td>$dN_{p}/dy$ [21]</td>
<td>1400</td>
<td>5100</td>
<td></td>
<td>$dE_t/dy$ (GeV)</td>
<td>2000</td>
<td>17000</td>
<td></td>
</tr>
<tr>
<td>$dN_{p}/dy$ [22]</td>
<td>80-120</td>
<td>1.5-5x10$^3$</td>
<td>$dE_t/dy$ (GeV)</td>
<td>200-300</td>
<td>5-15x10$^3$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In elementary proton-proton reactions, both the charged particle multiplicity $dN_{ch}/dy$ [14] and the average transverse momentum $<p_t>$ [15] grow only slowly (logarithmically) with $\sqrt{s}$. This would lead only to a modest growth in $dN_{ch}/dy$ (SPS:RHIC:LHC = 0.75:1:2.1), $<p_t>$ (0.92:1:1.35) and energy density (roughly the product $dN_{ch}/dy \cdot p_t$). However, also the scaling when going from pp to central AA reactions changes drastically over this energy range. While at low cms energy particle production is, to a very good approximation, proportional to the number of participants $N_{part}$, i.e. scales like $A$ in central AA collisions, hard processes, which dominate at high energies, scale proportional to the number of collisions $N_{col} (\propto A^{4/3})$. Compared to SPS, this leads to an additional enhancement factor for $dN/dy$ in central AA reactions of about three (assuming equal contributions of soft and hard processes) or six (assuming hard dominance) for RHIC and LHC, respectively. In addition, as structure functions rise strongly at small $x$, the number of partons available for collisions increases a function of energy. As an illustration, the number of partons inside the proton above 2 GeV is listed in the table as row ‘partons in p’ [16]. While some of this increase is reduced by shadowing in heavy nuclei, the net effect for particle production, roughly proportional to the square of parton number times shadowing reduction, is still very sizeable: A heavy nucleus contains about twice (four times) as many effective partons at RHIC (LHC) than at the SPS. The fact that nuclei get ‘denser’ at high energy is a great bonus which was not anticipated only a few years ago!

The right hand side of Table 1 summarises a selection of different measurements and estimates for global event features in Pb-Pb collisions (for a similar table, including some original references, see [17]). While $dN_{ch}/dy$ has been measured at the SPS and can be extrapolated with confidence from
the current RHIC results [18] to full RHIC energy, the predictions for LHC vary wildly between some 2000 and up to 8000 charged particles per unit rapidity. The main reasons for this spread are connected e.g. with uncertainties in gluon shadowing, final state parton saturation [19] and jet-quenching [20], all of which are expected to be rather different at RHIC and LHC. Therefore we might not be able to narrow down significantly this large range before LHC start-up even if more detailed results will become available from RHIC. Both the quantum mechanically determined 'formation time' of harder partons \( \propto 1/p_t \) and the increased interaction rate at high parton density decrease the thermalization time \( \tau^\text{QGP}_0 \) of the dense medium, leading to extreme energy densities \( \varepsilon(\tau_0) \propto 1/\tau_0 \) and temperatures at very early times during the collision at the LHC. As the system has to expand and cool to freeze-out conditions which are probably similar at all energies, QGP lifetime \( \tau^\text{QGP} \), freeze-out time \( \tau^\text{fo} \) and final (local) freeze-out volume \( V^\text{loc} \) likewise increase substantially. In particular the strong increase in QGP lifetime, both in absolute terms, in units of relaxation times (which in general decrease in denser systems), as well as compared to the hadronic phase, might be particularly relevant as it should enhance the relative contribution of signals generated in and sensitive to the QGP phase.

Finally, the last two rows show two representative pQCD calculations of initial parton density \( dN_p/dy \), dominantly gluons, and transverse energy production \( dE_t/dy \) [21,22]. While they differ in detail in their treatment of shadowing, parton saturation effects and \( p_t \) cut-off, they qualitatively agree that densities should increase by about an order of magnitude from RHIC to LHC, much stronger than the logarithmic rise seen in pp reactions.

In summary, by increasing the energy from soft to hard dominated regimes we gain independently on several fronts: More partons are available, their inelastic interactions get both more effective (already in pp) and in addition more frequent (in nuclei), and the thermalized high temperature phase is established more rapidly and lasts longer.

### 3.2. Qualitative changes at LHC

Besides better initial conditions, a number of qualitative changes are expected at high energy leading to a very different system and observables which are either new or can be studied with greater precision.

**Parton saturation:** At LHC, and to a smaller extend at RHIC, the density of low-\( x \) partons both in the incoming nuclei and in the final state after scattering are so high that the phase space for low momentum partons comes close to saturation and gluon merging becomes important to an extend that the reactions are sometimes referred to as a collision between two 'gluon walls' [19]. In the initial state, structure functions are modified by shadowing, and also in the final state the gluon density is reduced, limiting in a self-consistent way the infrared divergence of pQCD cross sections. Due to the large occupation numbers, the scattering process might be described by classical field theory, a new regime that becomes accessible for the first time with heavy ions at LHC. This 'gluon saturation' is expected to set in around 2 GeV at LHC [21] where pQCD might be applicable and theoretical concepts can be tested.

**Baryon density:** The incoming nuclei are slowed down during the collisions and the net baryon number is distributed over typically 3-5 units in rapidity. This leads to a high baryon density at SPS and below, whereas it is not sufficient to populate the central region at LHC, which will have a vanishing baryochemical potential \( \mu_b \) comparable to the early universe. Therefore very different regions of the phase diagram will be explored at low energy ('compression') and high energy ('heating'). At RHIC, first indications are that a baryon excess persists, which is however small and probably not very relevant \( (\mu_b/T \ll 1) \).
Energy density and pressure, as calculated on the lattice, are shown in Fig. 7, together with the estimated temperature range accessible at different machines. The SPS is certainly very close to, and likely already reaches beyond the hadronic phase; RHIC is ideally suited to explore the phase boundary in great detail; but only LHC will reach deep into the QGP and could approach the Stefan-Boltzman limit of an 'ideal gas' of QCD quanta. The region between $T_c$ and the SB limit is of particular interest in order to study the interactions responsible for the deviation from the ideal parton gas case.

**Hard processes:** Hard scattering with large momentum transfer becomes significant at RHIC (measurable up to about 20 GeV) and copious at LHC (up to several hundred GeV) [23]. These hard probes, in particular high $p_t$ jets and heavy quark production, are formed at very early times and therefore will test the surrounding dense medium. The study of the classical deconfinement signal, i.e. $J/\Psi$ suppression, can be extended at LHC to the $Y$ family. A full 'spectral analysis' of heavy quarkonia states [24] will be much more difficult at RHIC, as the cross sections are lower by at least an order of magnitude [22] and the initial temperature is likely to be below the melting point of the tightly bound $Y$ state (Fig. 8). Open charm (and eventually beauty) production will be sufficiently frequent to add these heavy quarks into the list available for abundance analysis ('chemical analysis' of parton or hadron ratios, currently limited to up, down and strange quarks). This will help in disentangling the different stages – initial production, pre-equilibrium and thermal processes, hadronization – which are currently a major source of confusion in interpreting the measured ratios. 'Jet-quenching', the medium induced energy loss of hard partons [20], is a new probe not available at the SPS. It might be visible already in the shape of the inclusive hadron $p_t$ spectrum at RHIC, but should greatly improve in terms of clarity and theoretical tractability once absolute jet cross sections and fragmentation functions are measured at LHC in the $E_t$ range of tens to hundreds of GeV.

**Thermal radiation:** Observation of direct photons emitted from the hot initial reaction volume is a crucial piece of information which is currently still lacking or at least ambiguous. As the signal strength improves dramatically with energy density and plasma life-time, it might be observable at the colliders, in particular if the background from high $p_t \pi^0$'s is reduced by thermalization and jet-quenching.
Event-by-Event physics: A unique feature of heavy ion collisions is the possibility to measure a number of observables on an even-by-event basis. Non-statistical fluctuations in these observables are of interest [25] because they are, in general, associated with critical phenomena in the vicinity of a phase transition, they can be related to thermal quantities (eg the heat capacity), or might be indicative of ‘anomalous’ events as suggested by cosmic ray observations (Centauro’s etc.). Currently, with clear guidance and precise predictions from theory missing, this subject is essentially data driven. However, as the accuracy will in general increase inversely with the number of observed particles \((1/\sqrt{N})\), EbE physics will become a precision instrument at high particle multiplicity and any observation of non-trivial fluctuating would be a clean ‘smoking gun’. Even in the absence of anomalous fluctuations, some EbE measurements will be extremely useful to study the correlation with other inclusive observables, like the azimuthal dependence of jet-quenching with respect to the event plane which can be determined via the elliptic flow.

3.3. Heavy ions in the LHC

The LHC will start to collide protons at \(\sqrt{s} = 14\) TeV in early 2006 and will provide the first heavy ion collisions (Pb-Pb) towards the end of its first year of operation at a total cms energy of 1148 TeV \((\sqrt{s} = 5.5\) TeV per nucleon for Pb-Pb). The sharing between proton and ion runs is expected to be similar to the one used in the past at the SPS, i.e. a long proton run \((10^7\) s effective time) followed by a few weeks \((10^6\) s) of heavy ion or proton-nucleus collisions. The ion luminosity is limited for heavy systems by the short beam life-time and energy deposition in the supra-conducting machine magnets (quench protection), and for light ions by the injector chain. It reaches values between \(3 \cdot 10^{31} cm^{-2} s^{-1}\) for light ions \((O^{16})\) and \(10^{27} cm^{-2} s^{-1}\) for Pb ions. Proton- or deuteron-nucleus collisions are feasible and foreseen, however only at equal magnetic rigidity per beam as the bending field is identical in both rings (‘two-in-one’ magnet design of the LHC). Further information on heavy ion operation of the LHC can be found in Ref. [26].

3.4. The dedicated heavy ion experiment ALICE

ALICE is a general-purpose heavy-ion detector designed to study the physics of strongly interacting matter and the quark-gluon plasma in nucleus-nucleus collisions at the LHC. It currently includes more than 900 physicists – both from nuclear and high energy physics – from about 70 institutions in 25 countries.

The detector is designed to cope with the highest particle multiplicities anticipated for Pb-Pb reactions \((dN/dy \approx 8000)\) and it will be operational at the start-up of the LHC. In addition to heavy systems, the ALICE Collaboration will study collisions of lower-mass ions, which are a means of varying the energy density, and protons (both pp and p-nucleus), which provide reference data for the nucleus-nucleus collisions.

ALICE (Fig. 9) consists of a central part, which measures hadrons, electrons and photons, and a forward spectrometer to measure muons. The central part, which covers polar angles from 45° to 135° over the full azimuth, is embedded in the large L3 solenoidal magnet. It consists of an inner tracking system (ITS) of high-resolution silicon tracking detectors, a cylindrical TPC, three particle identification arrays of time-of-flight (TOF), Ring Imaging Cerenkov (HMPID) and Transition Radiation (TRD) counters and a single-arm electromagnetic calorimeter (PHOS). The forward muon arm (2° – 9°) consists of a complex arrangement of absorbers, a large dipole magnet, and fourteen stations of tracking and triggering chambers. Several smaller detectors (ZDC, PMD, FMD, CASTOR, T0) are located at very forward angles.
Fig. 9: Artists view of the ALICE detector.

**Magnet:** The optimal choice for ALICE is a large solenoid with a weak field (<0.5 T) allowing full tracking and particle identification inside the magnet. The available space has to be sufficiently large to accommodate the PHOS, which must be placed at a distance of ≈ 5 m from the vertex, because of the large particle density. The magnet of the L3 experiment fulfils all these requirements.

**Inner Tracking System:** The basic functions of the inner tracker - secondary vertex reconstruction of hyperon and charm decays, particle identification and tracking of low-momentum particles, and improvement of the momentum resolution – are achieved with six barrels of high-resolution detectors. Because of the high particle density, the innermost four layers need to be truly two-dimensional devices, i.e. silicon pixel and silicon drift detectors. The outer layers, at r ≈ 50 cm, will be equipped with double-sided silicon micro-strip detectors. Four of the layers will have analog readout for independent particle identification via dE/dx in the non-relativistic region, which will give the inner tracking system a stand-alone capability as a low p_t particle spectrometer.

**Time Projection Chamber:** The need for efficient and robust tracking of up to 12000 charged particles within the acceptance has led to the choice of a TPC as the main tracking system. The inner radius of the TPC (r ≈ 90 cm) is given by the maximum acceptable hit density, the outer radius of 250 cm by the length required for a dE/dx resolution of <10%. The design of the readout chambers and electronics, as well as the choice of the operating gas, is optimised for good double-track resolution.

**Particle Identification System:** Particle identification over a large part of the phase space and for many different particles is an important design feature of ALICE. A large Time-of-Flight array (>100m^2) at a radius of about 3.5 m is made with novel multigap resistive plate chambers (MRPC) with an intrinsic time resolution of <100 ps. The proximity focusing RICH detector HMPID, of smaller acceptance and at larger radii, is optimised for the detection of high p_t particles and will extend the accessible momentum range for inclusive particle spectra into the semi-hard region. The six layer Transition Radiation Detector (TRD) will identify electrons with momenta above 1 GeV/c to study quarkonia suppression and heavy quark production (charm, beauty) in the central acceptance.

**Photon Spectrometer:** The electromagnetic calorimeter will be located below the interaction region at 4.6 m from the vertex, and covers 8 m^2 with 17k channels of scintillating PbWO_4 crystals to measure direct photons and high p_t neutral mesons. These very dense crystals are needed to cope with the large particle density, and to have sufficient light-output to allow readout with silicon photodiodes.
**Forward Muon Arm:** The forward muon arm is designed in order to cover the complete spectrum of heavy quark resonances, i.e., $J/\psi$, $\Psi'$, $Y$, $Y'$, $Y''$ (Fig. 8). It will measure the decay of these resonances, both in proton-proton and in heavy-ion collisions. The angular acceptance of the muon spectrometer is from $2^\circ$ to $9^\circ$ ($\eta = 2.5 - 4$). Its mass resolution will be around 100 MeV at 10 GeV, sufficient to separate all resonance states. It consists of a composite absorber, made with layers of both high- and low-Z materials, starting 90 cm from the vertex, a large dipole magnet with a 3 Tm field integral placed outside the L3 magnet, and 10 planes of thin, high-granularity tracking stations (Cathode Strip and Cathode Pad Chambers). A second absorber at the end of the spectrometer and four more detector planes (RPC’s) are used for muon identification and triggering. The spectrometer is shielded throughout its length by a dense absorber tube, of about 60 cm outer diameter, which surrounds the beam pipe.

**Forward Detectors:** ALICE uses a number of smaller detector systems (ZDC, PMD, FMD, CASTOR, T0) located at forward angles to define and trigger on global event characteristics. Four small and very dense calorimeters (Zero Degree Calorimeters, ZDC) are located about 100 m inside the machine tunnels on both sides of the interaction to define the impact parameter of the collision. A similar calorimeter (CASTOR), located closer to the interaction region on side of the experiment, will measure electromagnetic and hadronic transverse energy at large rapidity. The Forward Multiplicity Detector (FMD) measures charge particle production over a large fraction of phase space ($|\eta| < 4$). The Photon Multiplicity Detector (PMD) will search for non statistical fluctuations in the ratio of photons to charged particles, measure collective flow and transverse energy of neutral particles, and in addition determine the reaction plane. The T0 detector will measure the event time with great precision.

**Trigger, Data acquisition and Offline:** The ALICE trigger is foreseen to work on five levels, starting with a fast minimum bias interaction pre-trigger (issued after $< 1 \mu s$) to strobe some of the front-end electronics and ending with a high level online computing farm of several hundred PC’s (L3 trigger) which is intended for further event selection (for example on high $p_t$ jets or high mass lepton pairs) as well as online pre-tracking and event compression. Several detectors provide input to the different trigger levels to select e.g. for centrality, high $p_t$ electrons, muons, or photons.

The relatively short heavy-ion running period and the very large event sizes (up to 80 Mbyte even after zero suppression) determine the main features of the DAQ. In order to collect a sufficient number of events for physics analysis, the DAQ system has to be designed with a very large bandwidth up to 1.25 GByte/s on mass-storage. The DAQ architecture is based on a network of high-speed links linking all the data sources and the data destinations through a switch. This architecture provides the required flexibility and scalability to run in very different modes.

A new Off-line framework (ALIRoot) has been developed since 1998 based on all new C++ code and the OO paradigm. The ROOT framework was adopted as a base for this development, integrating currently the GEANT 3 and later also the GEANT 4 simulation package. At the moment a complete OO simulation of ALICE exists and the OO reconstruction code is being developed in this framework.

**4. SUMMARY**

The still very young field of ultra-relativistic heavy ion physics has proceeded since its inception in 1986 through three essential phases:

The initial round of ‘exploratory’ experiments has shown that appropriate detectors and analysis procedures can cope with the extreme particle densities produced in heavy ion collisions. They have qualitatively shown that an extended, interacting and very dense system has been formed that differs in many observables from the more elementary hadron-hadron reactions investigated in the past. Falling short of striking discoveries, this phase has nevertheless provided a ‘principle proof of feasibility’ and
has substantiated the expectation that heavy ion collisions are an appropriate tool to create equilibrium hadron matter and eventually the quark-gluon plasma.

The next phase was characterized by efforts to get a comprehensive and precise set of data and to come to a quantitative understanding of the experimental results. A close and very effective interaction between theory and experiment, models and data, has led to significant progress in identifying relevant ingredients and important microscopic processes.

The field is currently in its third, and most dramatic phase. Results from both AGS and SPS with really 'heavy' ion beams have produced puzzling results which strongly hint at a picture of high energy nuclear reactions almost too good to be true: i) a premordial phase of deconfined partons – the QGP? – responsible for quarkonium suppression, followed by ii) a transition regime with gradual onset of chiral symmetry breaking, leading to changes in the properties of light hadrons, concluded by iii) a gas of hadronic matter governed by the simple laws of thermodynamics. On the short term, further results from the fixed target program, where analysis and even new experiments are still under way, could provide us with additional and more complete data in order to substantiate (or refute) this scenario.

In the longer term, making use of RHIC and LHC for heavy ion collisions provides a unique opportunity for exploring the physics of QCD matter in a qualitatively very different region of extremely high energy density. RHIC and its four major experiments (STAR, PHENIX, BRAHMS and PHOBOS) are making the first step, starting in 2000 with Au+Au collisions at an energy an order of magnitude above what is currently available at the SPS.

Some 5 years from now, a new regime of very high energy density but low baryon density will be accessible with heavy ion collisions at the LHC. At some 30 times RHIC design energy, the step in energy from RHIC to LHC will be enormous, in fact larger than the one going from SPS to RHIC. LHC will reach, and even extend, the energy range probed by cosmic ray nucleus–nucleus collisions. Extrapolating from present results, all parameters relevant to the formation of the Quark–Gluon Plasma (QGP) will be more favourable: the energy density, the size and lifetime of the system, as well as relaxation times should all improve by a large factor, typically by an order of magnitude compared to SPS and even RICH. It should then be possible to obtain energy densities far above the deconfinement threshold, and to probe the QGP in its asymptotically free 'ideal gas' form. Unlike at lower energies, the central rapidity region will have a nearly vanishing baryon number density, similar to the state of the early universe. Reactions will be dominated in the early pre-equilibrium stage by a very dense system of semi-hard partons ('mini-jets'), which would lead to rapid thermalization and extremely high initial temperatures. Hard probes (heavy quark production, jets, even weak bosons) will be copiously produced well into the 100 GeV mass and momentum range providing a new and, with pQCD, well calibrated tool to study the QGP.

The LHC will therefore complement, extend and eventually succeed the ongoing heavy ion program by providing a very different, and in many instances significantly better, environment for the study of strongly interacting matter.

With the past at AGS/SPS very productive and the present at RHIC extremely exciting, the field of ultra-relativistic heavy ion physics may confidently look forward to a promising future at LHC.

214
References


INSTRUMENTATION

Olav Ullaland
CERN, Geneva, Switzerland

Abstract
Any device that is to detect a particle must interact with it in some way. If the particle is to pass through the detection system essentially undeviated, this interaction must be a soft electromagnetic one. There is another type of detector system that will measure the energy and the position of the particle by total absorption in these devices.

1. INTRODUCTION
This write-up of the Instrumentation lecture series given at the CERN-CLAF 2001 School of Physics, will give a general, although somewhat compressed, introduction to particle interaction with matter and magnetic fields. Tracking detectors and calorimeters will also be covered as well as particle identification systems. This write-up does not aim to be a complete stand-alone discussion of particle interactions nor to give an exhaustive coverage of experimental techniques in physics. This can be found in a number of excellent text books. Some of them are listed in reference [1]. The mathematical formalism that is used in this write-up is based on reference [2] and it is assumed that the reader is familiar with the formalism of electrodynamics.

The note will start out with a short review of particle interaction with fields and then we will discuss particle detection. At the end some common composite detection systems will be described.

2. PARTICLE INTERACTION WITH FIELDS
2.1. From Rutherford to Bethe-Bloch
The angular deflection of a particle with mass $m_1$, charge $eZ_1$ and velocity $v_0$ on a target particle of mass $m_2$ and charge $eZ_2$ can be written in the centre of mass coordinate system

$$\tan \frac{\Theta_{CM}}{2} = \frac{K}{v_0^2 b}$$

where

$$K = \frac{e^2 Z_1 Z_2}{4 \pi \varepsilon_0 m_r}$$

and

$$m_r = \frac{m_1 m_2}{m_1 + m_2}$$

and $b$ is the impact parameter as defined in Fig. 1. This is the classic nonrelativistic representation of

![Rutherford Scattering](image_url)

Fig. 1: Rutherford scattering
Rutherford scattering. By integrating over the impact parameter \( b \)

\[
\frac{dN}{d\Omega} = \frac{N_0}{256\pi^2\varepsilon_0^2}[nt]e^4\frac{Z_1^2Z_2^2}{(\frac{1}{2}m_1v_0^2)^2}\frac{1}{\sin^4\frac{\Theta_{CM}}{2}}
\]

where \( N_0 \) – number of beam particles, \( n \) – target material in atoms/volume and \( t \) – target thickness. As there is a screening by the electric field of the atom, there is a minimum scattering angle \( \Theta_{min} \) which is approximately inversely proportional to the momentum of the particle. For a single scattering, the mean scattering angle is thereby given as

\[
\overline{\Theta^2} \approx 2\Theta_{min}^2 \ln \frac{\Theta_{max}}{\Theta_{min}} \quad \text{or for } N \text{ multiple scatterings} \quad \overline{\Theta_{MS}^2} \approx \frac{N_0\rho dx}{A} \frac{2Z\alpha}{p/\beta c} \left\{ \ln() \right\}^2.
\]

The slowly varying logarithmic term will be ignored. Multiple Coulomb scattering can be treated as classic Rutherford scattering. If the energy of the particle is above the ionisation energy, \( I \), of the material, an electron escapes the atom. If the energy is below \( I \), no energy is transferred. The mean excitation energy is plotted in Fig. 2 where \( I/Z = 10 \pm 1 \text{ eV for } Z > 18 \). The nonrelativistic energy loss for a particle in matter can thereby be written as

\[
\frac{dE}{dx} = \frac{NZ_1^2e^4}{8\pi\varepsilon_0^2m_ev_0^2} \sum_{i=Z'}^{Z} \ln \frac{2m_i^2v_0^2}{m_eI_i}.
\]

If we ignore the slowly varying logarithmic term, the range, \( R \), of a particle in matter would be proportional to the squared of the ratio of the kinetic energy divided by the \( Z \) of the material. We can compare

![Fig. 2: Excitation energy divided by Z [3]](image-url)
this to the current wisdom of the Bethe-Bloch formula [3] plotted in Fig. 3:

\[
\frac{dE}{dx} = K Z^2 \frac{1}{A} \beta^2 \frac{1}{\beta^2} \ln \left( \frac{2m_e c^2 \beta^2 \gamma^2 T_{max}}{I^2} \right) - \beta^2 - \delta^2,
\]

\[
K = \frac{4\pi N_{A} r_e^2 m_e c^2}{A} = 0.307075 \text{ MeV g}^{-1}\text{cm}^2 \text{ for } A=1 \text{ g/mol and } T_{max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma \frac{m_e}{M} + \left(\frac{m_e}{M}\right)^2}.
\]

By comparing the relativistic and the nonrelativistic computation of the range, Fig. 4, we see that the range is approximately proportional to the energy square at low energy and proportional to the energy where \(dE/dx\) is about constant, on the Fermi plateaux. Further discussion can be found in [4].
2.2. Bremsstrahlung, photon pair production and radiation length

The Feynman diagrams for Bremsstrahlung and photon pair production are given in Fig. 5. Let us first consider Bremsstrahlung as a nonrelativistic radiative process. Let $b$ be the impact parameter. The peak electric field is proportional to $e/b^2$ and the characteristic frequency $\omega_c \propto \Delta t^{-1} \propto \frac{1}{b^2}$. This process is described by

$$dU = \int du = \int \left(\frac{dE}{d\omega}\right)^2 \cdot 2\pi b db = \int \frac{dU}{d\omega} d\omega \quad \Rightarrow \quad \frac{dU}{d\omega} = 4\alpha\beta e^2 \ln() .$$

The differential cross section for Bremsstrahlung can be written as

$$\frac{d\sigma_B}{d\omega} \approx Z^2 \frac{dN_\gamma}{d\omega} \sigma_T \quad N_\gamma \, : \, \text{photon density} \quad \sigma_T \, : \, \text{Thomson cross section}$$

$\downarrow$

$$\frac{dN_\gamma(\omega)}{d\omega} \approx \frac{2\alpha}{\pi \beta^2 \omega} [\ln()] \quad \sigma_T = \frac{8\pi}{3} (\alpha \lambda)^2$$

$$\sigma_B \sim 0.58 \, \text{mb} \cdot Z^2$$

By comparing the Feynman diagrams in Fig. 5, we would assume that the cross section for pair production would be very similar to that of Bremsstrahlung, $\sigma_{\text{pair}} = \frac{7}{9} \sigma_B \approx 0.45 \, \text{mb} \cdot Z^2$. The radiative energy loss by Bremsstrahlung can be written as

$$dE \approx \int_0^E \hbar \omega \frac{N_0}{A} d\sigma_B d\omega \quad \text{Define } X_0 \equiv \text{Radiative Mean Path} \quad \frac{1}{X_0} \equiv \frac{1}{E} \frac{dE}{\rho dx} \equiv \text{Radiation Length}$$

and

$$X_0^{-1} = \frac{16}{3} \frac{N_0}{A} Z^2 \alpha(\alpha \lambda)^2 [\ln()] \propto \frac{Z^2}{A} .$$

Radiation length as function of $Z$ is plotted in Fig. 6. The multiple scattering angle, $\Theta_{MS}^2$, can be expressed as function of the radiation length

$$\sqrt{\Theta_{MS}^2} = \frac{E_S}{p\beta} \sqrt{\frac{dx}{X_0}} = \frac{E_S}{p\beta} \sqrt{t} ,$$

where $t$ is the thickness of the material expressed in radiation lengths. We have here introduced the characteristic energy, $E_S \equiv m_e c^2 \sqrt{\frac{4\pi}{\alpha}} = 21.205 \, \text{MeV}$.

When we later will discuss calorimeters as particle detectors, we will expand the initiating particle in
ionisation. The classic way to do this is by using the Rossi II approximation to electromagnetic shower development [5].

1. Electrons lose a constant amount of energy, $\epsilon$, for each radiation length, $X_0$.
2. Radiation and pair production at all energies are described by the asymptotic formulae.

We will define the critical energy, $E_C$, as the energy where Bremsstrahlung is equal to ionisation. A high energy electron can lose energy by emitting gamma radiation if it is deflected. Below this critical energy of the material, the electron no longer loses energy in this way; it simply ionises atoms or is scattered. The fractional energy loss is plotted in Fig. 7 together with the critical energy as a function of $Z$. The photon total cross section is plotted in Fig. 8.

![Radiation length as function of Z](image)

**Fig. 6:** Radiation length as function of $Z$. The solid line is a fit to $A/Z^2$.

![Fractional energy loss and critical energy](image)

**Fig. 7:** a) Fractional energy loss per radiation length in lead as function of positron or electron energy [3] b) Critical energy as function of $Z$
2.3. Nuclear interaction length

The mean free path of a particle in a medium is a measure of its probability to undergo interactions of a given kind. It is related to the cross section corresponding to this type of interaction by the formula

$$\sigma \lambda = \frac{\Omega}{N} = \frac{A}{N_A \rho},$$

where $\sigma$ is the cross section [cm$^2$], $\lambda$ the mean free path [cm], $\Omega$ the volume of interaction, $N$ the number of target particles in $\Omega$, $A$ the atomic weight [g/mol], $N_A$ is Avogadro’s number and $\rho$ is the density [g/cm$^3$]. The relevant cross section is $\sigma_{\text{total}} - \sigma_{\text{elastic}} - \sigma_{\text{diffractive}}$. For hadronic interactions, the interesting quantity is the nuclear interaction length, $\lambda_I$. It is nearly energy independent and approximately proportional to $A^{1/3}$ as shown in Fig. 9.
2.4. Cherenkov radiation

The density effect in the energy loss, Fig. 3, is intimately connected to the coherent response of a medium to the passage of a relativistic particle. This causes the emission of Cherenkov radiation. Let us consider charged particle interaction with matter as described in Fig. 10.

If $\omega \ll \gamma m$ and $k \ll \beta \gamma m$, then

$$\cos \Theta_C = \frac{1}{\beta \sqrt{\varepsilon}} \iff \begin{cases} \omega = \beta \cdot \frac{k}{\varepsilon} & \text{from conservation of energy and momentum,} \\ \omega^2 - k^2 \varepsilon = 0 & \text{from dispersion relation.} \end{cases}$$

We have here considered the electric permeability, $\varepsilon$, as a real number. $\varepsilon$ for argon is plotted in Fig. 11. The refractive index of argon at NTP, can in the Sellmeier approximation be written as $(n - 1)^{10^6} = 0.05139 \cdot [16.88^2 - E^2]^{-1}$, where $E$ is the photon energy in eV. We observe that the refractive index is undefined at the value where the electric permeability goes from above 1 to below 1 and where the imaginary part becomes important. This is at the plasma frequency, $\omega_0$, for the material. The square of the plasma frequency is approximately proportional to the electron density in the material. By calculating the electromagnetic energy flow in a cylinder of radius $a$ around the track of the particle [2], we get after some steps

$$\left(\frac{dE}{dx}\right)_{b>a} \rightarrow \text{Re} \left[ \int_0^{\infty} \frac{z^2 e^2}{c^2} \left( -i \sqrt{\frac{\lambda^2 - \lambda}{\lambda}} \right) \omega \left( 1 - \frac{1}{\beta^2 \varepsilon(\omega)} \right) e^{-(\lambda + \lambda^*)a} d\omega \right].$$
where \( a \) is in the order of atomic dimension, \( b \) is the impact parameter and

\[
\lambda^2 = \frac{\omega^2}{v^2} - \frac{\omega^2}{c^2} \varepsilon(\omega) = \frac{\omega^2}{v^2} \left[ 1 - \beta^2 \varepsilon(\omega) \right].
\]

If \( \lambda \) has a positive real part, the integrand will vanish rapidly and all energy is deposited near the track. If \( \lambda \) is purely imaginary, the integrand is independent of \( a \) and some energy will escape as radiation. For \( \lambda \) to be purely imaginary, \( \beta^2 \varepsilon > 1 \), which is the Cherenkov condition. We also see that \( \frac{dE}{dx} \propto \sin^2 \Theta_C \). We will come back to detectors for Cherenkov radiation in Section 4.1.

### 2.5 Transition radiation

There is another type of radiation, transition radiation [8], that is emitted when a charged particle passes suddenly from one medium to another. If \( \varepsilon < 1 \) no real photon can be emitted for an infinitely long radiator. Due to diffraction broadening, there is a sub-threshold emission of real photons in thin radiators.

\[
\frac{d^2 S_0}{d\Theta d\omega} = \frac{2\alpha \hbar}{\pi \omega} \left[ \frac{1}{a_1} - \frac{1}{a_2} \right]^2 \quad \text{where} \quad a_i = \frac{1}{\gamma^2} + \Theta^2 + \frac{\omega_0^2}{\omega^2}.
\]

If \( \omega_2 \gg \omega_1 \) then the angle of maximum radiation, \( \Theta_{\text{max}} \approx \gamma^{-1} \), and the total radiated power \( S(\text{eV}) \approx 10^{-2} \cdot \gamma \). The classic way to increase the number of produced photons from transition radiation is to construct a periodic radiator as in Fig. 12. Further discussion can be found in [8]. We will come back to detectors for transition radiation in Section 4.1.
2.6. Magnetic fields and momentum measurements

The classic Lorentz force equation, $\vec{F} = q \cdot \vec{E} + q \cdot \vec{v} \times \vec{B}$, describes the force experienced by a charge $q$ in an environment with an electric field strength, $\vec{E}$, and a magnetic induction, $\vec{B}$. Magnet elements may be represented to a good approximation by a linear transformation matrix $M$ operating on the displacement-divergence vectors $(x, x')$ and $(y, y')$. The particle motion is along $z$ and the prime denotes $d/dz$ and $|M| = 1$. Fig. 13 gives a representation of ideal dipole and quadrupole magnets. By using the Lorentz force equation and the nomenclature from Fig. 13 a, a rectangular bending magnet can then be represented by

$$
\begin{bmatrix}
  x_2 \\
  x'_2
\end{bmatrix} =
\begin{bmatrix}
  \cos \beta & R \sin(\alpha + \beta) \\
  0 & \cos \alpha \cos \beta
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x'_1
\end{bmatrix} \rightarrow \begin{bmatrix}
  x_2 \\
  x'_2
\end{bmatrix} =
\begin{bmatrix}
  1 & R \sin(\phi) \\
  0 & 1
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x'_1
\end{bmatrix}
$$

and $\phi \approx 0.03BL/p$ with $B$ in kG, $L$ in m and $p$ in GeV/c. With the same considerations, we can now write the transformation matrix for an ideal quadrupole magnet as described in Fig. 13 b. Different matrices will apply in the focusing and the defocusing plane

$$
M_{\text{focusing}} =
\begin{bmatrix}
  \cos \omega d & \omega^{-1} \sin \omega d \\
  -\omega \sin \omega d & \cos \omega d
\end{bmatrix}
M_{\text{defocusing}} =
\begin{bmatrix}
  \cosh \omega d & \omega^{-1} \sinh \omega d \\
  \omega \sinh \omega d & \cosh \omega d
\end{bmatrix}
$$

where $d$ is the length of the element in m, $\omega^2 \approx 3k_p$ for $k$ in kG/cm and $p$ in GeV/c. In the same way as in optics, we can rewrite these matrices in a thin lens analogy. The effect of a thin lens is to change the divergence instantaneously without effecting the displacement. Each plane in a quadrupole corresponds to a thin lens of focal length $f$ and a drift length $s$.

$$
\begin{bmatrix}
  1 & s \\
  0 & 1
\end{bmatrix}
\begin{bmatrix}
  1 & 0 \\
  f^{-1} & 1
\end{bmatrix}
= 
\begin{bmatrix}
  1 + sf^{-1} & s(2 + sf^{-1}) \\
  f^{-1} & 1 + sf^{-1}
\end{bmatrix}
$$

Fig. 12: A periodic transition radiator, a, with coherent addition in point P. Figure b from [7] for $\gamma = 2 \cdot 10^4$, $l_1=25 \mu m$, $l_2=0.2$ mm and for polypropylene and air.
The situation normally considered when using thin lens approximation is when \( \omega d \to 0 \). We then get \( f^{-1} \to \pm \omega^2 d \) and \( s \to 0.5d \). It can be verified that a quadrupole doublet can have a net focusing/defocusing effect on a particle beam. Further reading can be found in [9].

The momentum of the charged particle is usually calculated from the curvature in a magnetic field. Consider a particle trajectory in a magnetic field as described in Fig. 14 a. The momentum of the particle is given by [10]

\[
p \approx \frac{q}{3} BR \sin \alpha \quad p \approx \frac{q}{3} \frac{BR_{\perp}}{\sin \alpha} \quad p \approx \frac{q}{3} \int \frac{B_{\perp} \, dl}{\Theta} \quad p \approx \frac{q}{3} B \left[ \frac{C^2 \sin \alpha}{2S} + \frac{S}{2 \sin \alpha} \right],
\]

where \( B \) is the magnetic induction, \( R \) is the radius of curvature in trajectory space, \( \alpha \) is the angle between the tangent to the trajectory and \( \vec{B} \), \( C \) is the length of the chord and \( S \) the related sagitta. \( \Theta \) is the bending angle. \( R_{\perp} \) is the radius of curvature of the trajectory projected in the plane perpendicular to \( \vec{B} \).

Fig. 13: Figure a) is a rectangular bending magnet. b) is for an ideal quadrupole configuration.

Fig. 14: a) A particle with charge \( q \) and momentum \( p \) in a uniform magnetic field \( \vec{B} \). b) Production of particle jets by 16 GeV negative pions in the first CERN liquid hydrogen bubble chamber. It was only 30 cm in diameter. CERN photo-ex/11465.
A beam with momentum-analysed particles may consist of particles of different mass. A standard device for velocity separation is generally a parallel plate capacitor of length $L$ along $z$. Consider a vertical, $y$, $E$ field. The deflection and the divergence difference between two particles with mass $m_1$ and $m_2$ can be written like

$$\Delta y = -\frac{1}{2}kEL^2 \quad \text{and} \quad \Delta y' = -kEL \quad \text{for} \quad k = \frac{e}{pc} \frac{\beta_1 - \beta_2}{\beta_1 \beta_2}.$$  

At high momenta, $k \sim \left(\frac{m_2^2 - m_1^2}{2}(pc)^2\right)$.  

2.7 Synchrotron radiation

It is well known that accelerated charges emit electromagnetic radiation. Deflection of ultra relativistic particles in magnetic fields leads to emission of radiation called synchrotron radiation. Following reference [2], and after some manipulation of the 4-vector potential caused by a charge in motion, it can be shown that

$$\vec{B} = \left[\vec{n} \times \vec{E}\right]_{\text{ret}} \quad \vec{E}(\vec{x}, t) = e \left[\frac{\vec{n} - \vec{\beta}}{\gamma^2 \left(1 - \vec{\beta} \cdot \vec{n}\right)^3 R^2}\right]_{\text{ret}} + \frac{e}{c} \left[\frac{\vec{n} \times \left[\left(\vec{n} - \vec{\beta}\right) \times \vec{\beta}'\right]}{1 - \vec{\beta} \cdot \vec{n}} \frac{R}{R^2}\right]_{\text{ret}},$$

where $d\tau = dt/\gamma$, $\vec{n}$ is the unit vector in the direction of $\vec{x} - \vec{r}(\tau)$, $\vec{\beta} = \vec{v}/c$, $\vec{\beta}' = d\vec{\beta}/dt$ and $R = x_0 - r_0(\tau_0)$ as defined in Fig. 15. The subscript $\text{ret}$ means that the quantity is to be evaluated at the retarded time $\tau_0$. The velocity field is essentially a static field that falls off as $R^2$, whereas the acceleration field depends linearly on $\vec{\beta}'$ and is inversely proportional to $R$. The total energy loss is given by

$$\frac{dW}{dt} = \frac{2c}{3}e^2 \beta^4 \gamma^4 \frac{1}{r^2}$$

and the root mean square of the emission angle is $\left[\Theta^2\right]^{1/2} \approx \frac{1}{\gamma}$

and is independent of the vector relationship between $\vec{\beta}$ and $\vec{\beta}'$. $r$ is the bending radius of the charged particle. As $\beta \gamma = p/m$, synchrotron radiation is mainly observed with low-mass particles.
Fig. 16: a) Straw tube design for charged particle detection. b) Primary and secondary ion pair production per unit length for gases at atmospheric pressure and for minimum ionising particles [11].

3. PARTICLE DETECTION

3.1. Gaseous Detectors

Consider the simple detector geometry given in Fig. 16a. Let the metallic anode wire have a radius \( r_0 \) and the metallic cathode cylinder a radius \( R \), \( r_0 \ll R \). A minimum ionising particle will leave along the track in the gas, \( N_{\text{total}} \) ion pairs and free electrons. \( N_{\text{total}} \approx 5 \cdot Z/\text{cm} \) at NTP, where \( Z \) is the atomic number of the filling gas as given in Fig. 16b. The electric field next to the anode can be written as \( E_0 = V_0 \left/ \left[ r_0 \ln \left( \frac{R}{r_0} \right) \right] \right. \). \( V_0 \) is the voltage difference between the anode and the cathode. Let \( \alpha^{-1} \) be the mean free path between each ionisation induced by the free electrons on the gas atoms as the electrons fall towards the anode. \( \alpha = \alpha(E) \) and \( E = E(r) \). \( \alpha \) is generally known as the First Townsend coefficient. The total amount of electrons reaching the anode, the gas amplification constant, can then be written as

\[
M = e^{\int_{r_0}^R \alpha(r) dr} \quad \text{Korff approximation: } \alpha = A p e^{-B p/E} \quad \rightarrow \quad M = \exp \left[ A \frac{V_0}{B \ln \left( \frac{R}{r_0} \right) e \frac{B p_0}{V_0} \frac{R}{r_0}} \right]
\]

where \( p \) is the gas pressure and \( A \) and \( B \) are gas dependent constants, see Table 1. Fig. 17a shows an example of the characteristic energy for an electron as function of the electric field. Clearly the field has to be very high for the energy to be above the ionisation potential. A dramatic change can be observed in a gas mixture, Fig. 17b. This is the Penning effect. The metastable states are responsible for this effect. It allows to convert a greater fraction of the energy which was initially lost in excitation, into ion pairs, if the admixed gas component has a lower ionisation potential than the available energy of the excited...
Fig. 17: a) Characteristic energy and drift velocity of electrons in CO\(_2\) as function of the electric field [11]. b) The first Townsend coefficient, \(\alpha\), as function of the electric field in a gas mixture of Ne and Ar [12].
Table 1: Parameters for the Korff approximation for the first Townsend coefficient $\alpha$ [11].

<table>
<thead>
<tr>
<th>Gas</th>
<th>A $\text{cm}^{-1}\text{Torr}$</th>
<th>B $\text{Vcm}^{-1}\text{Torr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>Ne</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Ar</td>
<td>14</td>
<td>180</td>
</tr>
<tr>
<td>Xe</td>
<td>26</td>
<td>350</td>
</tr>
<tr>
<td>CO$_2$</td>
<td>20</td>
<td>466</td>
</tr>
</tbody>
</table>

Fig. 18: Absorption coefficient for water as function of wavelength.

State. The excited atoms of the principal gas are then able to transfer this energy in ionising collisions with molecules of the admixed component. It is thereby possible to have $\alpha \gg 0$ for reasonable electric fields and to begin the gas amplification process and to start the electron avalanche towards the anode. The main bulk of the electrons will be produced within a few anode diameters from the anode.

At the same time as the electrons are created, positively charged ions are created. The de-excitation of these ions will normally involve the emission of a photon. If the photon energy is above the ionisation threshold for other molecules in the set-up, new free electrons will be created and thereby permanent discharges. Poly-atomic gases are therefore added to noble gases as quenchers $^1$. An example is shown in Fig. 18. The different operating regions for a chamber are shown in Fig. 19 for a heavily ionising $\alpha$ particle and for a $\beta$ particle which is near minimum ionising. There will also be effects due to the way the electrons are collected at the anode. The electric field of the chamber will be screened by the positive ions. The gas amplification will therefore change as the angle between the electric field and the ionising particle changes. The drift velocity and the diffusion of the electrons will change with the gases that are used in the chamber. A magnetic field will change the drift path of the electrons as well as the diffusion.

$^1$Energy can be stored in molecules as translational, rotational and vibrational energy. Translation can occur in the $x$, $y$ or $z$ direction. Rotation can occur around the $x$, $y$ or $z$ axis, except for linear molecules which only have two axes or rotation. Vibrations involve movements of the atoms of a molecule which produce no net translation or rotation. These various movements are a result of the combination of the normal modes of vibration. For a triatomic molecule these normal modes are symmetric, asymmetric and bending vibrations. In symmetric vibration, the two bonds shorten and lengthen together. In asymmetric vibration, one bond shortens while the other lengthens. In bending vibration, it is the bond angle that oscillates. From: http://www4.ncsu.edu:8030/ajschult/pchem/water/water.html
The schematics of a classic multiwire proportional chamber is shown in Fig. 20 together with the electric field configuration. The relationship between the voltage, the charge and the magnitude of the electric field can be written as

\[ V_s(z) \approx \frac{2\pi l}{s} - \ln \left[ 4 \sin^2 \left( \frac{\pi x}{s} \right) + 4 \sinh^2 \left( \frac{\pi y}{s} \right) \right] \]

which gives

\[ Q = \frac{V_0}{\frac{2\pi l}{s} - 2 \ln \frac{\pi d}{s}} \]

and

\[ E_0 = \frac{sV_0}{\frac{\pi d}{2} \left[ l - \frac{s}{\pi} \ln \frac{\pi d}{s} \right]} \].

The positive pulses that are induced by the positive ions onto the neighbouring anode wires are much greater than the negative pulses induced electrostatically. The net effect is therefore positive. The signal induced is mainly by the positive ions moving in the high electric field. Following reference [11] and assuming that all charges are created a distance \( \lambda \) from the anode, we can write

\[ V_{electron} = -\frac{Q}{ICV_0} \int_{r_0}^{r_0+\lambda} \frac{dV}{dr} dr = -\frac{Q}{2\pi \varepsilon_0 l} \ln \frac{r_0 + \lambda}{r_0}, \]

\[ V_{ion} = +\frac{Q}{ICV_0} \int_{r_0+\lambda}^{R} \frac{dV}{dr} dr = +\frac{Q}{2\pi \varepsilon_0 l} \ln \frac{R}{r_0 + \lambda}, \]

where \( IC \) is the total capacitance. As \( \lambda \) is in the order of a few \( \mu m \), \( V_{electron} \sim V_{ion}/100 \).

The simple detector configurations that we have discussed above, can be readily understood. For a more complicated geometry, computer aided calculations will be necessary [13]. All derivative devices like the Drift Chamber (DC), the Time Projection Chamber (TPC), the MicroStrip Gas Chamber (MSGC), the Gas Electron Multiplier (GEM) or similar detectors, will adhere to the same principles as we have discussed above. Further reading can be found in [11,14]. Reference [12] gives a thorough discussion of electrons and ions in gases.

Fig. 19: Gain and voltage characteristics for \( \alpha \) and \( \beta \) particles in a proportional counter [11].
In the discussion of the Bethe-Bloch formula in page 219 and plotted in Fig. 3, the energy loss of a charged particle in matter is a function of the $\beta\gamma$ of the particle. We normally define it into four regions: the $1/\beta^2$ region, the minimum ionising region, the relativistic rise and the Fermi plateau region. As the total number of ions created in the gas is proportional to the energy lost by the charged particle, we can use the charge collected in the proportional chamber together with a measurement of the momentum of the particle as a measure for the energy of the particle. Thereby we can make a mass hypothesis for the particle. The measured charge has unfortunately large statistical fluctuations due to the primary ionisation process and the gas amplification process. In addition, there is the creation of $\delta$ electrons, knock-on energetic electrons which will travel far in a gas volume, that will add to the uncertainty of the measurement. The charge collected in a thin gas gap is well described by a Landau-like function [15]. It can be approximated to

$$\Psi(\lambda) = \sqrt{\frac{e^{-(\lambda + e^{-\lambda})}}{2\pi}}$$

where $\lambda = R(E - E_p)$. $E_p$ is the most probable energy loss and $R$ is a detector specific constant. A histogram of the charge collected in a proportional chamber is given in Fig. 21 a.
To suppress large fluctuations, it is normal to take many samples along the track and to make a truncated mean measurement. That is, to ignore the $x\%$ highest and $y\%$ lowest measurements and take the mean of the remainder. This estimate will be near Gaussian distributed. An example of the result is shown in Fig. 21 b.

Fig. 21: a) Ionisation energy loss for 3 GeV/c $\pi$ in Ar/CH₄. Also indicated is the average and the most probable energy loss. b) Results from the DELPHI experiment at LEP. Notice the clear electron band and the proton and deuterium bands which are well populated at low momenta.

3.2. Solid State Detectors

The resistivity of a material falls into one of three classes. The insulator with a volume resistivity between $10^9$ and $10^{20}$ $\Omega$cm, the semiconductor between $10^{-2}$ and $10^1$ $\Omega$cm and the conductor between $10^{-6}$ and $10^{-3}$ $\Omega$cm. Each electron in a solid has a certain total energy which is made up from kinetic and potential energy. The complete energy axis can be divided into the forbidden bands, no electron can have these energies, and the allowed bands where there may be electrons. The band structure is shown in Fig. 22 a. The distinction between an insulator and a semiconductor is one of degree rather than kind. The insulator has a band gap between the valence band and the conduction band of $\geq 3$ eV, whereas this band gap is between 0.1 and 2.5 eV in a semiconductor. The density of states, $S(W)$, as function of the energy, $W$, 
Fig. 22: a) Allowed and forbidden bands. In a metal one band is only partially filled. The valence band is (nearly) full in a semiconductor and the conduction band is (nearly) empty. The energy gap is larger between the valence band and the conduction band in an insulator than in a semiconductor. b) The number of states, $S(W)$, and the probability that the states are occupied, $P(W)$, in a semiconductor. c) Excitation of electrons in a semiconductor due to the passage of a charged particle.

is of course zero in the band gaps. For the lower edge of the conduction band

$$S(W) = \frac{4\pi}{h^3} (2m_e)^{\frac{3}{2}} (W - W_c)^{\frac{1}{2}}$$

and for the upper edge of the valence band

$$S(W) = \frac{4\pi}{h^3} (2m_h)^{\frac{3}{2}} (W_v - W)^{\frac{1}{2}}$$

where $W_i$ is the energy at the edge of the band and $m_i$ denotes the effective masses of carriers, holes or electrons, at the two band edges. The curves for $S(W)$ are to a first approximation parabolic. Let $P(W)$ describe the probability that a state at an energy $W$ is occupied by an electron. These states will follow the Fermi-Dirac statistics and for a system in thermal equilibrium, we get

$$P(W) = \frac{1}{1 + e^{\frac{W - W_F}{kT}}}$$

$W_F$ is the Fermi level. It is defined as the energy where $P(W) = 1/2$. The number of electrons occupying states at different energies, is then given by $S(W) \cdot P(W)$. See Fig. 22b. By integrating this function from $W_c$, the lower edge of the conduction band, to $W_i$, the upper edge of the band, we get the number of electrons, $n_i$, in the conduction band. $W_i$ is set equal to $\infty$ for ease of computation. The number of holes, $p$, in the valence band can be found in the same way by using $1 - P(W)$ instead of $P(W)$ and integrating from $W_v$ at the top of the valence band. We then get for the Fermi level for a thermal equilibrium situation

$$W_F = \frac{W_c + W_v}{2} - \frac{\kappa T}{2} \ln \frac{p}{n} - \frac{3}{4} \frac{\kappa T \ln \frac{m_e}{m_h}}{\kappa T \ln \frac{m_e}{m_h}}.$$

When $p \neq n$, then the Fermi level shifts towards the band with the majority carrier. In the case when $p$ and $n$ are controlled by an external device, the system may be far from equilibrium and $pn \neq n^2$.

One way to move the Fermi level is to dope the semiconductor. There are two kinds of dopant, donors and acceptors. A list is given in Table 2. Assume a n-type doped Si. There are 5 electrons in
Table 2: A list of a number of pairs of terms that go with donors or acceptors [16].

<table>
<thead>
<tr>
<th>Type of dopant</th>
<th>Donor</th>
<th>Acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semiconductor type</td>
<td>n-type</td>
<td>p-type</td>
</tr>
<tr>
<td>Majority carrier</td>
<td>electrons</td>
<td>holes</td>
</tr>
<tr>
<td>They occur in the conduction band</td>
<td>valence band</td>
<td></td>
</tr>
<tr>
<td>Minority carrier</td>
<td>holes</td>
<td>electrons</td>
</tr>
<tr>
<td>Possible dopant in Si or Ge</td>
<td>P, As, Sb</td>
<td>B, Al, Ga</td>
</tr>
<tr>
<td>Dopant in group</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Charge in ionised dopant</td>
<td>+ve</td>
<td>-ve</td>
</tr>
<tr>
<td>Fermi level is nearer</td>
<td>conduction band</td>
<td>valence band</td>
</tr>
</tbody>
</table>

the M-shell of the dopant. One electron, the conduction electron, has a binding energy of 10 to 50 meV. The Fermi level is very near the conduction band. For a p-type doped Si, the configuration is different. There are 3 electrons in the M-shell. One electron is missing, i.e. a hole. The Fermi level is very near to the valence band. Let us examine a p-n junction. From the considerations above, we can see that the net current of electrons and holes across the junction must be zero and that $pn = n^2$ at all points. From the cancellation of the current

$$-n(x)e\mu_e \frac{dV(x)}{dx} + eD_e \frac{dn(x)}{dx} = 0$$

or

$$\mu_e \frac{dV(x)}{dx} = D_e \frac{1}{n} \frac{dn(x)}{dx},$$

where $V(x)$ is the potential, $n(x)$ is the electron density at a distance $x$ from the junction, $\mu_e$ is the electron mobility and $D_e$ is the electron coefficient of diffusion, $D = D_0 \exp\left(-\frac{\Delta W}{\kappa T}\right)$ and $\Delta W$ is the energy barrier. By integrating the current equation from well inside the n-side to well inside the p-side, we get

$$V_b = V_n - V_p = \frac{D_e}{\mu_e} \ln \frac{n_n}{n_p}.$$

This potential difference is called the barrier potential. The barrier potential is generally lower than the band gap. In a state of equilibrium, the Fermi level is constant throughout the system as shown in Fig. 23. From an inspection of the voltage across the junction and integrating over the full length, $X$, of the depletion length, we see that

$$X = \sqrt{\frac{2eV}{\varepsilon N_d}}$$

and the capacitance

$$C(V) = \varepsilon A \frac{X}{X},$$

where $N_d$ is the number of donors and $A$ is the area of the junction. If an external voltage is applied across the junction, the system is no longer in equilibrium and the Fermi level is no longer constant throughout the system. With a forward bias, $V$ is reduced and thereby $X$ is reduced as the bias voltage is subtracted from the barrier potential. A reverse bias adds to the barrier potential and the result is therefore a wider depletion layer.

The mobility and conductivity in a semiconductor is shown in Fig. 24. The volume resistivity for Ge is $\leq 0.49 \ \Omega \text{m}$ at 300K whereas it is of the order of 100 $\Omega \text{m}$ in Si due to impurities. This will correspond for a Si crystal of $100 \times 100 \times 0.3 \ \text{mm}^3$ to $\sim 4 - 5 \times 10^8$ free carriers. A minimum ionising particle would, in comparison, create $\sim 3 - 4 \times 10^4$ e-h pairs. To reduce the number of free carriers, one can either deplete the crystal or do a freeze-out as shown in Fig. 24. Depletion by applying a reverse bias is the most commonly used technique for particle detection. A possible lay-out is shown in Fig. 25.
Fig. 23: A p-n junction without bias. Note the peak electric field at the boundary between the p and the n. Also note the clear depletion layer shown in the last sketch. [16].
Depending of the detector configuration and the electronics that are used, space resolution of some 10 \( \mu \text{m} \) can be achieved. An introduction to semiconductors can be found in references [16,17].

![Diagram of semiconductor properties](image)

**Fig. 24:** The variation with temperature of majority and minority carrier concentration, mobility \( \mu \) and conductivity \( \sigma \) for a semiconductor.

![Diagram of silicon detector configuration](image)

**Fig. 25:** Sketch of possible configurations for a silicon detector.

### 3.3 Scintillators

A charged particle traversing matter leaves behind it a wake of ionisation, excited molecules and molecules in metastable states. Certain types of molecules will release a small fraction (~ 3%) of the absorbed energy as optical photons. Scintillating materials fall into one of two classes, organic or inorganic. An excited molecule can lose energy by vibration, dissociation or a resonance transfer to another molecule. It can also lose by emitting a photon. This is the scintillation or fluorescence. Fluorescence is
prompt emission of photons as a result of an absorption of energy. The decay time of this photon emission is in the range from ns to $\mu$s. Let us first consider a perfect crystal structure as shown in Fig. 22c. The lifetime of a free hole is in the range of $10^{-12}$ s. This recombination may lead to photon emission, but it is too fast to be observed. The free electron might be trapped to the hole. This system is called an exciton. It is free to move through the crystal. The exciton can emit light when hitting an activator center and transferring their binding energy to the activator levels, which will subsequently de-excite. Impurities or activators like Thallium, are therefore generally added to the crystal. In an imperfect crystal, there are lattice defects and impurities as shown in Fig. 26a. The excitation energy can here be dissipated not only by luminescence, but also by thermal dissipation, quenching, and by metastable levels, traps. The metastable systems will lose the energy either via thermal or vibrational dissipation or via a radiationless transition. The photon emission by luminescence can be written like

$$\frac{dL}{dx} = \frac{A}{1 + B \frac{dE}{dx}} \frac{dE}{dx}$$

or for small $dE/dx$

$$\frac{dL}{dx} \propto \frac{dE}{dx}.$$

Table 3 gives selected physical properties of some inorganic crystalline scintillators. Note the fairly strong dependence on the light output as function of the temperature. Also the decay time is effected by the temperature. The light intensity as function of the wavelength is plotted in Fig. 26b.

Table 3: Physical properties of some inorganic scintillators. BGO is bismuth germanate and PWO is lead tungstate.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>BGO</th>
<th>CsI:Tl</th>
<th>CsI</th>
<th>PWO</th>
<th>NaI:Tl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (g/cm³)</td>
<td>7.13</td>
<td>4.53</td>
<td>4.53</td>
<td>8.26</td>
<td>3.67</td>
</tr>
<tr>
<td>Radiation length (cm)</td>
<td>1.12</td>
<td>1.85</td>
<td>1.85</td>
<td>0.89</td>
<td>2.59</td>
</tr>
<tr>
<td>Wave length (nm)</td>
<td>480</td>
<td>565</td>
<td>310</td>
<td>420</td>
<td>410</td>
</tr>
<tr>
<td>Light yield ($L$) (% of NaI)</td>
<td>10</td>
<td>85</td>
<td>7</td>
<td>0.2</td>
<td>100</td>
</tr>
<tr>
<td>Decay time (ns)</td>
<td>300</td>
<td>1000</td>
<td>6+35</td>
<td>5+15+100</td>
<td>250</td>
</tr>
<tr>
<td>$dL/dT$ (%/°C at 18°C)</td>
<td>-1.6</td>
<td>0.3</td>
<td>-0.6</td>
<td>-1.9</td>
<td>0</td>
</tr>
<tr>
<td>Refractive index</td>
<td>2.15</td>
<td>1.8</td>
<td>1.8</td>
<td>2.29</td>
<td>1.85</td>
</tr>
</tbody>
</table>

The behaviour of organic scintillators are about the same as for inorganic. They are mainly based on benzene, $C_6H_6$. The p-electrons\(^2\) forms the basis for the scintillation mechanism. They are quantised in series of singlets, $S_{ij}$, and triplets, $T_{ij}$, as in Fig. 27. A practical organic scintillator uses a solvent for the primary scintillation and then large concentration of a primary Fluor and smaller amount of secondary Fluor to shift down the final photon spectrum as indicated in Fig. 28a. The difference in wavelength between absorbed and emitted quanta, the Stokes shift, is shown in Fig. 28b. This shifting of the primary photon spectrum from a decay time of $\leq 10^{-9}$ s and a peak wave length of $\sim 300$ nm, is essential for an efficient photon detection as is discussed in Section 3.4. Theory and application are well covered in reference [19].

\(^2\)Some bonds in benzene [18]: Benzene is a planar molecule with six C-C bond distances of equal length. The observed bond distance (1.40 angstroms) is midway between the $sp^2$-$sp^2$ single-bond distance (1.46 angstroms) and $sp^2$-$sp^2$ double-bond distance (1.34 angstroms) seen in conjugated dienes and is consistent with the bond order of 1.5 predicted by resonance theory. (Bond order is an index of bond strength. A bond order of 1 indicates that a single s bond exists between two atoms, and a bond order of 2 indicates the presence of one s and one p bond between two atoms. Fractional bond orders are possible for resonance structures, as in the case of benzene.) Benzene is a regular hexagon; all bond angles are 120°.
Fig. 26: a) shows the band structure of an inorganic crystalline scintillator. b) scintillation intensity as function of wavelength.
3.4. Photon Detection

Photon detection around the visible spectrum is generally built around two processes, the photon to electron conversion, i.e. the photocathode, and the electron multiplication to generate a readable output signal. The operation of a photocathode is based on the photoelectric effect. For a semiconductor this may be considered as a three step process. In the first step, the photoelectron is excited from the valence band to the vacuum level by the absorption of the light quanta. The second step is to transport the excited electron through the semiconductor film to the semiconductor/vacuum interface. The third step is an escape over the surface barrier into the vacuum. The efficiency of the photoelectric emission is determined by the efficiency in each step. At the surface of a semiconductor there is always some bending of the energy bands due to the presence of donors and acceptors with energy levels in the forbidden zone. See Section 3.2. for discussion of the energy bands. p-type materials are therefore the most efficient photoemitters. These are semiconductors like (Na$_2$KSb)Cs, Cs$_3$Sb and GaAs:Cs·O. The emission energy for the photoelectron excited by a monochromatic light, will range from zero to a maximum given by $E_{\text{light}} - [E_{\text{electron affinity}} + E_{\text{gap width}}]$. In addition to photoemission, the photocathodes will also produce thermionic emission. This will present itself as an unwanted background. A summary of photocathode characteristics is given in Table 4.

Table 4: Summary of photocathode characteristics [20]. $E_g$ is the energy gap between valence and conduction band. $E_A$ is the electron affinity or the energy gap between the conduction band and the vacuum level.

<table>
<thead>
<tr>
<th>Cathode</th>
<th>Peak $\lambda$ ($\AA$)</th>
<th>Maximum Quantum Yield at peak (%)</th>
<th>$\lambda_0$ ($\AA$) [at 1% of peak]</th>
<th>Maximum (µA/lm)</th>
<th>Typical Thermionic Emission (A/cm$^2$)</th>
<th>$E_g$ (eV)</th>
<th>$E_A$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag·O-Cs</td>
<td>8000</td>
<td>0.5</td>
<td>$\approx$ 12000</td>
<td>60</td>
<td>$10^{-12}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cs$_3$Sb on MnO</td>
<td>4000</td>
<td>20</td>
<td>6500</td>
<td>80</td>
<td>$10^{-15}$</td>
<td>1.6</td>
<td>0.45</td>
</tr>
<tr>
<td>(Cs)Na$_2$KSb</td>
<td>4000</td>
<td>30</td>
<td>9500</td>
<td>550</td>
<td>$10^{-15}$</td>
<td>1.0</td>
<td>0.35</td>
</tr>
<tr>
<td>K$_2$CsSb</td>
<td>4000</td>
<td>30</td>
<td>6600</td>
<td>100</td>
<td>$&lt;10^{-17}$</td>
<td>1.0</td>
<td>0.90</td>
</tr>
<tr>
<td>Na$_2$KSB</td>
<td>4000</td>
<td>30</td>
<td>6600</td>
<td>130</td>
<td>$&lt;10^{-17}$</td>
<td>1.0</td>
<td>0.75</td>
</tr>
<tr>
<td>GaAs:Cs·O</td>
<td>7800</td>
<td>&gt; 20</td>
<td>9500</td>
<td>&gt; 1000</td>
<td>$&lt;10^{-14}$</td>
<td>1.4</td>
<td>&lt; 0</td>
</tr>
</tbody>
</table>
Fig. 28: a) shows the wavelength shifting in an organic scintillator. b) The emitted wavelength is always longer or equal to the incident wavelength. The difference is absorbed as heat in the atomic lattice of the material.

Table 5: Physical properties of some possible gaseous photon converters [21]. TEA is Triethyl amine and TMAE is Tetrakis(dimethylamino)ethylen. See also Fig. 29 b.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Ionisation energy (eV)</th>
<th>$p_0$ at 300K (Torr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>10.4</td>
<td>47</td>
</tr>
<tr>
<td>Acetone</td>
<td>9.69</td>
<td>250.7</td>
</tr>
<tr>
<td>Benzene</td>
<td>9.24</td>
<td>101.8</td>
</tr>
<tr>
<td>TEA</td>
<td>7.5</td>
<td>73.2</td>
</tr>
<tr>
<td>TMAE</td>
<td>5.36</td>
<td>0.50</td>
</tr>
</tbody>
</table>

These photocathodes are very sensitive to air and will easily oxidise. They must therefore be deposited and operated in ultra high vacuum. There is another type of photocathodes that can be operated in an inert gas. They are mainly used as an additive to the gaseous detectors. Some properties of these components are listed in Table 5. CsI, an alkali halide crystal, which falls into the first category, is used successfully as a solid photon converter in gaseous detectors. Quantum efficiency for some photocathodes are plotted in Fig. 29.

The electron amplification process in a gaseous detector is described in 3.1.. In a vacuum tube, the secondary-electron process is used. The process of secondary-electron emission is similar to photomission. It is different for metals than for insulators or semiconductors. When a clean metal surface is exposed to an energetic primary electron, the secondary-electron emission takes place by raising the energy level in the conduction band to the one of the vacuum level. One primary electron may have sufficient energy to excite one or more secondary electrons. Clean metals are poor secondary-electron emitters whereas insulators and semiconductors are efficient. Clearly the electron affinity plays a crucial role and an efficient secondary-electron emitter should have an $E_A$ zero or negative.
Fig. 29:  a) Quantum efficiency as function of wavelength for semitransparent photocathodes. Plot from http://www.hamamatsu.com/. b) Quantum efficiency as function of the photon energy for some photosensitive vapours. Repotted from [21].
From the discussion above, a photon detector is therefore either a gaseous detector where a photosensitive vapour has been added to the gas or a vacuum tube with a photocathode and electron multiplication based on secondary-electron emission. The microchannel plate is an example of the latter. A microchannel plate is a two dimensional array of microchannels as shown in Fig. 30 a. The microchannel itself has an internal diameter of 8 to 45 µm and is lined by a resistive secondary emission film. The schematics of a photoelectron multiplier tube is shown in Fig. 30 b. Dynodes of secondary-electron emitters are placed in the voltage chain between the photocathode and the anode. The total current gain is equal to the product of the current gain at each dynode. The production of secondary electrons is Poissonian distributed. The fluctuations on the final number is therefore mainly induced at the first dynodes where the number of primary electrons are small.

![Fig. 30: a) Operation of a single microchannel. Photocathode not shown. b) Photoelectron multiplier tube with photocathode and dynodes of secondary-electron emitting material. c) Is a sketch of a hybrid photo diode.](image)

The novel hybrid photo diodes [22] is a vacuum tube with a semitransparent photocathode, a focusing accelerating field and a silicon sensor at the anode, see Fig. 30c. The primary photoelectron will have a kinetic energy of 10 to 20 keV when it impinges onto the sensor and will produce in the order of 4 to 5000 electron-hole pairs. This ensures a very good signal resolution. For further reading on photon detectors, references [20,21] are recommended.

4. DETECTION SYSTEMS

Charged particle detection is already covered in Sections 3.1., 3.2. and 3.3.. We will in this Section look into some composite systems and also include detection of uncharged particles.

4.1. Particle Identification

The condition for Cherenkov radiation was given on page 223. The Cherenkov photons of wavelength $\lambda$ are emitted under an angle $\Theta_C$, where $\Theta_C$ is given by $\cos \Theta_C = 1/ \left[ \beta n(\lambda) \right]$ and the number of photons emitted by

$$\frac{dN}{d\lambda} = 2\pi \alpha \left( \frac{d}{\lambda} \right)^2 L \sin^2 \Theta_C,$$

$L$ is the length, $n(\lambda)$ is the refractive index of the radiator and $\alpha = 1/137$ is the fine-structure constant. The total number of photons emitted between $\lambda_1$ and $\lambda_2$ by a singly charged particle is then given like

$$N_{\lambda_1 \rightarrow \lambda_2} = 4.6 \times 10^6 \left[ \frac{1}{\lambda_2(\text{Å})} - \frac{1}{\lambda_1(\text{Å})} \right] L(\text{cm}) \sin^2 \Theta_C.$$
The number of detected photons is equal to the total number of emitted photons multiplied by the quantum efficiency of the detection system. If the momentum $p$ of the particle is determined as discussed on page 226, the mass, $m$, of the particle is given by $m = p/\beta\gamma$. The uncertainty of the mass measurement is given by

$$\frac{\Delta m}{m} = \sqrt{\left(\frac{\Delta p}{p}\right)^2 + (\gamma^2 \tan \Theta_C \Delta \Theta_C)^2}.$$  

We can therefore use the measurement of the Cherenkov angle and the knowledge on the momentum of the particle to make mass hypothesis for the charged particle.

Cherenkov detectors consist of a radiator with a refractive index $n$ and a photon detector. They are frequently categorised into two classes, the threshold detector and the ring imaging detector. The threshold detector will discriminate between particles where $\beta n > 1$ and particles below this threshold. They are mainly used in monochromatic beams to discriminate between particles of different masses. The ring imaging detector [23] exploits simultaneously more or less all the properties of Cherenkov radiation. That is, the existence of a threshold, the dependence of the number of photons, the dependence on the charge of the particle and the dependence of the Cherenkov angle. A practical detector will need to have a high efficiency for single photon detection and have a high space resolution. Cherenkov detectors are sketched in Fig. 31.

![Cherenkov Detector Diagrams](image)

Fig. 31: A threshold Cherenkov detector is outlined in a). b) shows a sketch of the principles of operation of a ring imaging Cherenkov detector. c) is an event display of a simulated event including a $B^0 \rightarrow \pi^+\pi^-$ event on the photodetector plane of a ring imaging Cherenkov detector. The detected photons are indicated by dots [24].

In the same way as we have done for the Cherenkov radiation, we can construct a detection system for transition radiation. Transition radiation is discussed on page 224. We recall that the total radiated power in eV is about 1% of $\gamma$ of the particle and that the angle of maximum radiation is very forward peaked as it is inversely proportional to $\gamma$. The energy of the emitted radiation has to be much larger than the plasma frequency of the radiator material and will in general be between 1 and 10 keV. A practical transition radiation detector will therefore consist of a radiator with many interfaces or foils and a detector for soft x-rays. The foils will be of low Z-material in order to limit the scattering and the detector could be proportional chamber based on Xenon for a maximum conversion efficiency. The number of foils is a trade-off between production of x-rays and the absorption of them. The detection of electrons from the converted x-ray is also made difficult by the fact that they are very near to the trajectory of the particle and therefore the ionisation caused by the particle in the gas of the detector. A high energy particle will create $\sim 200$ electron-ion pairs per cm in Xe gas at NTP. The conversion of a 5 keV x-ray will create $\sim 400$ electron-ion pairs. The detector is therefore usually thin in order to enhance the discrimination power of the detector. One finds in general [25] that

$$\frac{\text{ionisation} + \text{detected transition radiation}}{\text{ionisation}} \sim Z^{-\frac{7}{2}}.$$  

Transition radiation detectors are used for $e$ to $\pi$ separation.
Particle identification by measuring the energy loss by the charged particle in matter, is discussed in Section 3.1. We will here only recall that this technique generally calls for a rather thick gas volume where many samples of the energy loss are taken along the track. As for the other identification methods, it is assumed that the momentum of the particle is determined by other means.

Particle identification can also be done by measuring the time difference between detection points along the particle track. This is the Time-of-Flight, TOF, technique. It is similar to the Cherenkov detection as it is a pure measurement of the $\beta$ of the particle. A TOF measurement can in principle be used for any value of $\beta$. A TOF system will typically be a large hodoscope of scintillators coupled to a photomultiplier read-out system for a measurement of the impact time, $T_1$, at this distance from the vertex, together with a knowledge of the time $T_0$. $T_0$ is the creation time of the event or the bunch crossing of the particle beams in the experiment. It could also be a smaller hodoscope placed very near to the vertex of the event. For certain applications, it is also possible to forego the $T_0$ by assuming that one of the particles detected in the far hodoscope has a velocity $\beta \simeq 1$. With the measurement of the time difference and the track length, $\beta$ of the particle is known. A small uncertainty on the time measurement is required in order to keep the track length inside reasonable limits and thereby limit the size of the hodoscope. Other fast detectors can be used. Some of them are discussed in [26,27]. A summary on particle identification is given in Table 6.

Table 6: Table from [25]. It was published in 1980, just 10 years after the first publication by G. Charpak on proportional chambers and 3 years after the first publication of J. Séguinot and T. Ypsilantis on ring imaging Cherenkov detectors. 1974 was the year of the first publication on drift chambers by G. Charpak and collaborators. References to updated numbers has been added as a last column to the table.

<table>
<thead>
<tr>
<th>Technique</th>
<th>$\gamma$ range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple ionisation measurements in homogeneous medium</td>
<td>$\gamma &lt; 6$</td>
<td>As $dE/dx \propto 1/\beta^2$ for $\gamma \leq 6$, powerful discrimination obtainable with modest resolution in the $dE/dx$ measurement. [4]</td>
</tr>
<tr>
<td>Time-of-Flight</td>
<td>$\gamma \leq 6$</td>
<td>$\sigma \leq 300$ ns has been achieved with large scintillator systems. $\sigma \leq 100$ ns with special spark chambers [27]</td>
</tr>
<tr>
<td>Gas threshold Cherenkov</td>
<td>$\gamma &gt; 10$</td>
<td>Not suitable for storage ring application where $4\pi$ coverage is required. [21]</td>
</tr>
<tr>
<td>Multiple ionisation measurements on relativistic rise in homogeneous medium</td>
<td>$2 \leq \gamma \leq 50$</td>
<td>Requires $\sigma(dE/dx) \sim 2 – 3%$; achieved so far in planar geometry only, but not in storage ring application. [4]</td>
</tr>
<tr>
<td>Imaging Cherenkovs</td>
<td>$2 &lt; \gamma \leq 200$</td>
<td>Cherenkov photons detected with UV-sensitised proportional chamber structures; or special optical systems, with severely limited phase-space acceptance for particles. [23]</td>
</tr>
<tr>
<td>Transition radiation</td>
<td>$\gamma \geq 1000$</td>
<td>Useful as compact threshold detector for specialised applications. [8]</td>
</tr>
</tbody>
</table>

4.2. Calorimeters

A class of detectors measures the energy and the position of particles through total absorption in these detectors. In the process of absorption, secondary particles are generated. These secondary particles will as well react inside the detector. The process gives rise to a shower. The growth and decay of a shower is a tremendously complex statistical process where several physical mechanisms participate. The net result is however that the initiating particle is expanded in ionisation and ultimately in heat. The temperature gradient and the shock wave accompanying can be measurable. We will discuss the atomic excitation
and ionisation which is created in the shower process and leave the calorimetric effect here. There are several reasons why the calorimeters are attractive. They are the only way to detect and measure neutral particles and as the absorption of the energy of the particle is a statistic process, the precision of the energy measurement of the particle should therefore be proportional to $1/\sqrt{E}$. From the discussion of the momentum measurement on page 226, we see that $\sigma_p/\sqrt{p^2} \propto \sigma_{R\phi}/(L^2B)$ in a spherical coordinate system. A calorimetric measurement can therefore become more accurate than a measurement of the curvature of the particle track in a magnetic field.

Calorimeters are normally classified into two main groups. The electromagnetic and the hadronic calorimeters. They are, as the name indicate, optimised for measuring particles that interact electromagnetically or has strong interactions with the absorbing material. Subclass of each are the sampling calorimeters and the homogeneous calorimeters. Fig. 32 a shows the development of electromagnetic showers. Fig. 32 b and c shows sketches of these detectors.

Electron pair production and Bremsstrahlung are the dominant interaction process for high energy photons and electrons. From Fig. 7 a, we see that Bremsstrahlung is nearly independent of energy above $\sim 1$ GeV. The same is the case for pair production which is plotted in Fig. 8. The secondary particles in the electromagnetic shower are electrons, positrons and photons and they will interact with the matter in the same way as the parent particle. The number of particles in the shower will therefore increase until the average energy of the particles is below a threshold to stop further particle creation. The shower will then spend its energy in ionisation or Compton scattering and will slowly die off. The change of the relative strength of the interaction processes are characterised by the critical energy, $E_C$, of the material. (See the discussion on page 221 and Fig. 7 b.) The lateral development of the shower is mainly governed by the electrons that do not radiate, but have enough energy to travel far away from the axis. From this rather sketchy discussion of the electromagnetic shower which follows the Rossi II approximation on page 221, we see that the natural unit to express the shower in, is in units of radiation length, $X_0$. The shower maximum will be different for electrons and for photons. Experiments have shown that for an initiating particle of energy $E$, the shower maximum in units of $X_0$ is

\[ E \text{ in } X_0 = 3 \]  

From [18]: Calorimeters have been designed in great variety. One type in widespread use, called a bomb calorimeter, basically consists of an enclosure in which the reaction takes place, surrounded by a liquid, such as water, that absorbs the heat of the reaction and thus increases in temperature. Bomb calorimeters have been developed to the point that heats of combustion of organic materials can be measured with results reproducible within 0.01 percent.
\[ t(X_0) = \ln \frac{E}{E_C} - 1.0 \quad \text{for electrons} \quad t(X_0) = \ln \frac{E}{E_C} - 0.3 \quad \text{for photons} . \]

The lateral shower spread is expressed in Molière units. It is defined as

\[ R_M = \frac{\text{Characteristic Energy} \cdot \text{Radiation Length}}{\text{Critical Energy}} = 21 \left[ \text{MeV} \right] \frac{X_0}{E_C} \propto \frac{A}{Z} \left( \frac{g}{cm^2} \right) \]

\( \sim 95\% \) of the shower is contained inside a cone with radius \( 2R_M \). The longitudinal track length, which is proportional to the energy deposit, is well approximated to

\[ T = \mathcal{F}(z) \frac{E}{E_C} X_0 \quad \text{where} \quad \mathcal{F}(z) \simeq e^z \left[ 1 + z \ln \left( \frac{z}{1.526} \right) \right] \quad \text{and} \quad z = 4.58 \frac{Z}{A} \frac{E_{\text{cut-off}}}{E_C} . \]

The cut-off energy, \( E_{\text{cut-off}} \), is the minimum energy of a particle that can be detected in the calorimeter.

For a sampling electromagnetic calorimeter as sketched in Fig. 32b, let us assume that we sample \( N \) times the energy deposited by the shower. Let \( x \) be the distance along the shower axis between each sample. Let us furthermore assume that each sample is statistically independent. For large \( N \), the energy resolution can then be written like

\[ \frac{\sigma(E)}{E} = 3.2\% \sqrt{\frac{E_C}{\mathcal{F}}} \sqrt{\frac{x}{X_0}} \sqrt{\frac{1}{E}} . \]

The development of a hadronic shower is much more complex than the electromagnetic processes. It is mainly propagated by hadronic interactions. At high energies, these processes are characterised by multi particle production and particle emissions from nuclear disintegration of excited nuclei. Due to the relatively frequent generation of \( \pi^0 \)-s, there is also an electromagnetic component present in hadronic showers. A sizeable amount of the available energy is converted into excitation and break-up of nuclei. Only a small fraction of this energy will eventually appear as detectable signal and with large event-to-event fluctuations. The hadronic multiplication process is measured by the nuclear interaction length, \( \lambda_I \), which is nearly energy independent. \( \lambda_I \) is practically proportional to \( A^{1/3} \) as shown in Fig. 9. Some characteristics for high energy hadronic showers may be written like

Shower maximum:

\[ l_{\text{max}} \approx \left[ 0.6 \log(E) - 0.2 \right] \lambda_I , \quad \text{with} \ E \ \text{in GeV} . \]

Shower depth for 95\% longitudinal containment:

\[ l_{95\%} \approx l_{\text{max}} + 4E^{0.15} \lambda_I , \quad \text{with} \ E \ \text{in GeV} . \]

Contrary to electromagnetic showers, which develop in sub-nanosecond time, the physics of hadronic showers is characterised by different time scales, the slowest of which, de-excitation of heavy nuclei, may reach a microsecond. Massive hadron calorimeters are commonly used for the detection of neutrinos, as the secondaries from weak interactions are mostly hadrons.

In a sampling calorimeter, Fig. 32b, the particle absorption and shower sampling is separated. A fraction of the total energy is sampled in the active detector. The active detector can be any of the detectors that we have discussed above, like scintillators, ionisation chambers, wire chambers or solid state detectors. The particle absorbers in an electromagnetic calorimeters will be of low \( X_0 \) material like lead. A hadronic calorimeter will use material like iron, with long \( \lambda_I \). The energy resolution for a large sampling electromagnetic calorimeter will be in the range of 10\% at 1 GeV. For large volume sampling hadronic calorimeters, it will be in the order of 80\%.
The reason for this moderate energy resolution in a hadron calorimeter, is the large fluctuations as a fair fraction of the incident energy is spent in back-scattering, energy leakage by muons, neutrinos and slow neutrons. In addition, a large fraction is lost by nuclear excitation and nucleonic evaporation. All these processes do not give an observable signal in the detector. This can be demonstrated by measuring the ratio of a signal from an electron and a signal from a hadron. This is the electron-hadron ratio, $e/h$. If $e/h = 1.0$, the calorimeter is said to be compensating. If it differs from unity by more than 5% or 10%, detector performance is compromised because of fluctuations in the $\pi^0$ content of the cascades. In most cases $e/h$ is greater than unity, particularly if little hydrogen is present. Energy resolution of $30 – 40\% / \sqrt{E}$ has been demonstrated for a well compensated hadronic sampling calorimeter [25,29,30].

In a fully active and homogeneous electromagnetic calorimeter as shown in Fig. 32c, the energy is typically measured by collecting photons as these detectors are normally made from low $X_0$ scintillators as in Table 3. In other material like lead glass, 55% PbO and 45% SiO$_2$, the light emission is by Cherenkov radiation. The energy resolution is limited by photon statistics. If we take a scintillating crystal, which is virtually free from intrinsic fluctuations, we get

$$\sigma(E) / E = \frac{1}{\sqrt{N_{pe}}} = \frac{1}{\sqrt{E(\text{GeV})N_{pe}/\text{GeV}}} ,$$

where $N_{pe}/\text{GeV}$ is the observed number of photons per energy unit. To get this number, the absolute light yield, the number of emitted photons for each energy unit, has to be multiplied with the light collection efficiency, the geometrical efficiency of the photon detector and the quantum efficiency integrated over the emission spectrum. In addition, there will be effects from lateral shower leakage, the punch-through and the material in front of the detector. $\sigma(E) / E$ in lead glass will be in the range of $\geq 5\% E^{-\frac{1}{2}}$ whereas for NaI(Tl) it will be in the range of $\geq 1.5\% E^{-\frac{3}{4}}$ for $E$ in GeV.

For real calorimeters in real experiments, the energy resolution is usually written as

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus \frac{c}{E} ,$$

where $a$ is the stochastic term. $b$ is a constant term which normally describes the overall inter-calibration. $c$ is the contribution from the electronics noise. Energy resolution for some homogeneous electromagnetic calorimeters in large experiments is given in Fig. 33. We see that the energy resolution for the BGO calorimeter in this experiment is well described by the parameters: $a \approx 1.9 \times 10^{-2}$, $b \approx 5 \times 10^{-3}$ and $c \leq 10^{-4}$. About 5400 crystals constituted the calorimeter. The final overall resolution compares well to the test beam result, which was given as $\sigma(E) / E = \left[1.54 / \sqrt{E} + 0.38\right] \%$.

![Fig. 33: Energy resolution as function of energy for large detector systems. Replotted from [32].](image)

The complex computation of shower development is well described in simulation programs like [31]. Reference [30] discuss the physics of electromagnetic and hadronic calorimeters.
5. CONCLUSION

We have in this write-up looked into the more general considerations of particle detection and left out many aspects that make a detector a success or a failure. Electronics is surely a major component. Some good books and talks on this subject are listed in reference [33]. Trigger/DAQ and Data Analysis are taken care of in separate contributions to this school. We have, though, left out all the magic, the witchcraft and we will not even mentioned the Hopi Snake dance which has proven most useful in cases of high voltage break-down. The Ouija board becomes handy when the answer can not be found on the Web and no trace of the book listed in the references can be found in the library. At no point was all the hard work which is the essential component to a working detector discussed. Obnoxious things like detector ageing has been happily swept under the carpet together with the most valuable detector control systems. It can nonetheless be a useful precaution to check the detector performance against Ohm’s law if the fuses keeps popping and against Newton’s laws of motion if the detector keeps falling down.

Disclaimer
The numbers, formulae, figures and references are believed to be correct, but are not guaranteed to be so.

and Acknowledgement
At the end, I will thank all my collaborators and in particular Erich Albrecht, Tito Bellunato, Ariella Cattai, Carmelo D’Ambrosio, Martyn Davenport, Thierry Gys, Christian Joram, Wolfgang Klempt, Martin Laub, Georg Lenzen, Dietrich Liko and Niko Neufeld for their help to make both the lectures and the write-up. I am indebted to the Organising Committee for inviting me to this very inspiring school. First and foremost I will though, thank all the students of the school for the stimulating discussions and the thought-provoking questions. I also hope you have had as much sweat, tears and fun reading this as I have had in writing it.
References


M. Alemi Nucl. Phys. Proc. Suppl. B78 (1999);


The online version of the FLUKA manual (fluka99.manual)
http://fluka.web.cern.ch/fluka/material/Fluka/head.html


TRIGGER AND DATA ACQUISITION

Nick Ellis
EP Division, CERN, 1211 Geneva 23, Switzerland

1. INTRODUCTION

These lectures concentrate on experiments at high-energy particle colliders, especially the general-purpose experiments at the Large Hadron Collider (LHC). These represent a very challenging case that illustrates well the problems that have to be addressed in state-of-the-art high-energy physics (HEP) trigger and data-acquisition (T/DAQ) systems. It is also where the author is working (on the trigger for the ATLAS experiment at LHC) and so is the example that he knows best. However, the lectures start with a more general discussion, building up to some examples from LEP that had complementary challenges to LHC and is a good reference point to see how HEP T/DAQ systems have evolved in the last few years.

Students at this school come from various backgrounds – phenomenology, experimental data analysis in running experiments, and preparing for future experiments (including working on T/DAQ systems in some cases). These lectures try to strike a balance between making the presentation accessible to all, and going into some details for those working more directly with T/DAQ.

1.1 Definition and scope of trigger and data-acquisition

T/DAQ is the online system that selects particle interactions that are potentially of interest for physics analysis (trigger), and that takes care of collecting the corresponding data from the detectors, putting them into a suitable format and recording them on permanent storage (DAQ). Special modes of operation need to be considered, e.g. the need to calibrate different detectors in parallel outside of normal data-taking periods. T/DAQ is often taken to include associated tasks, e.g. run control, monitoring, clock distribution and book-keeping, all of which are essential for efficient collection of the data and for its subsequent offline analysis.

1.2 Basic trigger requirements

As introduced above, the trigger is responsible for selecting interactions that are potentially of interest for physics analysis. These interactions should be selected with high efficiency, the efficiency should be precisely known (since it enters in the calculation of cross-sections), and there should not be biases that affect the physics results. At the same time, a large reduction of rate from unwanted high-rate processes may be needed to match the capabilities of the DAQ system and the offline computing system. High-rate processes that need to be rejected may be instrumental backgrounds or high-rate physics processes that are not relevant for the analyses that one wants to make. The trigger system must also be affordable, which implies limited computing power. As a consequence, algorithms that need to be executed at high rate must be fast. Note that it is not always easy to achieve the above requirements (high efficiency for signal, strong background rejection and fast algorithms) simultaneously.

Trigger systems typically select events\(^1\) according to a “trigger menu”, i.e. a list of selection criteria – an event is selected if one or more of the criteria are met. Different criteria may correspond to different signatures for the same physics process – redundant selections lead to high selection efficiency and allow the efficiency of the trigger to be measured from the data. Different criteria may also reflect the wish to concurrently select events for a wide range of physics studies – HEP “experiments” (especially those with large general-purpose “detectors” or, more precisely, detector systems) are really experimental facilities. Note that the menu has to cover the physics channels to be studied, plus

\(^1\) The term “event” will be defined in Section 3 – for now, it may be taken to mean the record of an interaction.

253
additional data samples required to complete the analysis (e.g. measure backgrounds, and check the detector calibration and alignment).

1.3 Basic data-acquisition requirements

The DAQ system is responsible for the collection of data from detector digitisation systems, storing the data pending the trigger decision, and recording data from the selected events in a suitable format. In doing so it must avoid corruption or loss of data, and it must introduce as little dead-time as possible (“dead-time” refers to periods when interesting interactions cannot be selected – see below). The DAQ system must, of course, also be affordable which, for example, places limitations on the amount of data that can be read out from the detectors.

2. DESIGN OF A TRIGGER AND DATA-ACQUISITION SYSTEM

In the following a very simple example is used to illustrate some of the main issues for designing a T/DAQ system. An attempt is made to omit all the detail and concentrate only on the essentials – examples from real experiments will be discussed later.

Before proceeding to the issue of T/DAQ system design, the concept of dead-time, which will be an important element in what follows, is introduced. “Dead-time” is generally defined as the fraction or percentage of total time where valid interactions could not be recorded for various reasons. For example, there is typically a minimum period between triggers – after each trigger the experiment is dead for a short time.

Dead-time can arise from a number of sources, with a typical total of up to \( O(10\%) \). Sources include readout and trigger dead-time which are addressed in detail below, operational dead-time (e.g. time to start/stop data-taking runs), T/DAQ down time (e.g. following a computer failure), and detector down time (e.g. following a high-voltage trip). Given the huge investment in the accelerators and the detectors for a modern HEP experiment, it is clearly very important to keep dead-time to a minimum.

In the following, the design issues for a T/DAQ system are illustrated using a very simple example. Consider an experiment that makes a time-of-flight measurement using a scintillation-counter telescope, read out with time-to-digital converters (TDCs), as shown in Fig. 1. Each plane of the telescope is viewed by a photo-multiplier tube (PMT) and the resulting electronic signal is passed to a “discriminator” circuit that gives a digital pulse with a sharp leading edge when a charged particle passes through the detector. The leading edge of the pulse appears a fixed time after the particle traverses the counter. (The PMTs and discriminators are not shown in the figure.)

Two of the telescope planes are mounted close together, while the third is located a considerable distance downstream giving a measurable flight time that can be used to determine the particle’s velocity. The trigger is formed by requiring a coincidence (logical AND) of the signals from the first two planes, avoiding triggers due to random noise in the photo-multipliers – it is very unlikely for there to be noise pulses simultaneously from both PMTs. The time of arrival of the particle at the three telescope planes is measured, relative to the trigger signal, using three channels of a TDC. The pulses from each of the three planes going to the TDC have to be delayed so that the trigger signal, used to start the TDC (analogous to starting a stop-watch), gets there first.

The trigger signal is also sent to the DAQ computer, informing it to initiate the readout. Not shown in the figure is logic that prevents further triggers until the data from the TDC have been read out into the computer – the so-called dead-time logic.
2.1 Traditional approach to trigger and data-acquisition

The following discussion starts by presenting a “traditional” approach to T/DAQ (as might be implemented using, for example, NIM and CAMAC electronics modules\(^2\), plus a DAQ computer). Note that this approach is still widely used in small test set-ups. The limitations of this model are described and ways of improving on it are presented. Of course, a big HEP experiment has an enormous number of sensor channels (up to \(O(10^8)\) at LHC), compared to just three in the example. However, the principles are the same, as will be shown later.

Limitations of the T/DAQ system shown in Fig. 1 are as follows:

1. The trigger decision has to be made very quickly because the TDCs require a “start” signal that arrives before the signals that are to be digitised (a TDC module is essentially a multi-channel digital stop watch). The situation is similar with traditional analogue-to-digital converters (ADCs) that digitise the magnitude of a signal arriving during a “gate” period, e.g. the electric charge in an analogue pulse – the gate has to start before the pulse arrives.

2. The readout of the TDCs by the computer may be quite slow, which implies a significant dead-time if the trigger rate is high. This limitation becomes much more important in larger systems where many channels have to be read out for each event. For example, if 1000 channels have to be read out with a readout time of 1 \(\mu\)s per channel (as in CAMAC), the readout time per event is 1 ms which excludes trigger rates above 1 kHz.

![Diagram of a simple experiment with its T/DAQ system](image)

Figure 1: Example of a simple experiment with its T/DAQ system

The “readout model” of this traditional approach to T/DAQ is illustrated in Fig. 2, which shows the sequence of actions – arrival of the trigger, arrival of the detector signals (followed by digitisation and storage in a data register in the TDC), and readout into the DAQ computer. Since no new trigger can be accepted until the readout is complete, the readout dead-time is given by the product of the trigger rate and the readout time.

\(^2\) NIM and CAMAC modules are electronic modules that conform to agreed standards – modules for many functions needed in a T/DAQ system are available commercially.
2.2 Local buffer

The traditional approach described above can be improved by adding a local “buffer” memory into which the data are moved rapidly following a trigger, as illustrated in Fig. 3. This fast readout reduces the dead-time, which is now given by the product of the trigger rate and the local readout time. This approach is particularly useful in large systems where the transfer of data can proceed in parallel to many local buffers (e.g. one local buffer for each crate of electronics) – local readout can remain fast even in a large system. Also, the data may be moved more quickly into the local buffer within the crate than into the DAQ computer.

The addition of a local buffer reduces the effective readout time, but the requirement of a fast trigger still remains. Signals have to be delayed until the trigger decision is available at the digitisers. This is not easy to achieve, even with very simple trigger logic – typically one relies on using fast (air-core) cables for trigger signals with the shortest possible routing, so that the trigger signals arrive before the rest of the signals (which follow a longer routing on slower cables). It is not possible to apply complex selection criteria on this time-scale.

2.3 Multi-level triggers

It is not always possible to simultaneously meet the physics requirements (high efficiency, high background rejection) and achieve an extremely short trigger “latency” (time to form the trigger decision and distribute it to the digitizers). A solution is to introduce the concept of multi-level triggers,
where the first level has a short latency and maintains high efficiency, but only has a modest rejection power. Further background rejection comes from the higher trigger levels that can be slower. Sometimes the very fast first stage of the trigger is called the “pre-trigger” – it may be sufficient to signal the presence of minimal activity in the detectors at this stage.

The use of a pre-trigger is illustrated in Fig. 4. Here the pre-trigger is used to provide the start signal to the TDCs (and to gate ADCs, etc.), while the main trigger (which can come later and can therefore be based on more complex calculations) is used to initiate the read out. In cases where the pre-trigger is not confirmed by the main trigger, a “fast clear” is used to re-activate the digitizers (TDCs, ADCs, etc.).

![Diagram](image)

**Figure 4: Readout system with pre-trigger and fast clear**

Using a pre-trigger (but without using a local buffer for now), the dead-time has two components. Following each pre-trigger there is a dead period until the trigger or fast clear is issued (defined here as the trigger latency). For the subset of pre-triggers that give rise to a trigger, there is an additional dead period given by the readout time. Hence, the total dead-time is given by the product of the pre-trigger rate and the trigger latency, added to the product of the trigger rate and the readout time.

The two ingredients — use of a local buffer and use of a pre-trigger with fast clear — can be combined as shown in Fig. 5, further reducing the dead-time. Here the total dead-time is given by the product of the pre-trigger rate and the trigger latency, added to the product of the trigger rate and the local readout time.

### 2.4 Further improvements

The idea of multi-level triggers can be extended beyond having two levels (pre-trigger and main trigger). One can have a series of trigger levels that progressively reduce the rate. The efficiency for the desired physics must be kept high at all levels. The initial levels can have modest rejection power, but they must be fast since they see a high input rate. The final levels must have strong rejection power, but they can be slower because they see much lower rate (thanks to the rejection from the earlier levels).

In a multi-level trigger system, the total dead-time can be written as the sum of two parts, the trigger dead-time summed over trigger levels, and the readout dead-time. For a system with \( N \) levels, this can be written:

\[
\left( \sum_{i=2}^{N} R_{i-1} \times L_i \right) + R_{N} \times T_{\text{R}}
\]
where \( R_j \) is the rate after the \( j^{th} \) trigger level, \( L_j \) is the latency of the \( j^{th} \) trigger level, and \( T_{\text{LRO}} \) is the local readout time. Note that \( R_j \) corresponds to the post-trigger rate.

In the above, two implicit assumptions have been made: (1) that all trigger levels are completed before the readout starts, and (2) that the pre-trigger (i.e. the lowest-level trigger) is available by the time the first signals from the detector arrive at the digitizers. The first assumption results in a long dead period for some events – those that survive the first (fast) levels of selection. The dead-time can be reduced by moving the data into intermediate storage after the initial stages of trigger selection, after which further low-level triggers can be accepted (in parallel with the execution of the later stages of trigger selection on the first event). The second assumption is hard to satisfy in practice, but as discussed below the requirement can be avoided, e.g. in collider experiments with bunched beams.

In the next section, aspects of particle colliders that affect the design of T/DAQ systems are introduced. Afterwards, the discussion returns to readout models and dead-time considering the example of LEP experiments.

![Readout system diagram](image)

**Figure 5:** Readout system using both pre-trigger and local buffer

### 3. COLLIDER EXPERIMENTS

In high-energy particle colliders (HERA, LEP, LHC, Tevatron), the particles in the counter-rotating beams are bunched. Bunches of particles cross at regular intervals and interactions only occur during the bunch crossings. Here, the trigger has the job of selecting the bunch crossings of interest for physics analysis, i.e. those containing interactions of interest.

In these notes, the term “event” is used to refer to the record of all the products from a given bunch crossing (plus any activity from other bunch crossings that gets recorded along with this). Be aware (and beware!) – the term “event” is not uniquely defined! Some people use the term “event” for the products of a single interaction between incident particles. Note that people sometimes unwittingly use “event” interchangeably to mean different things.

In e\(\bar{e}\) colliders, the interaction rate is very small compared to the bunch-crossing rate (because of the low e\(\bar{e}\) cross-section). Generally, selected events contain just one interaction – i.e. the event is generally a single interaction. This was the case at LEP and is also true at the e–p collider HERA. In contrast, at the LHC with the design luminosity of \(10^{34} \text{ cm}^{-2}\text{s}^{-1}\), each bunch crossing will contain on average about 25 interactions. This means that an interaction of interest, e.g. one that produced \(H \rightarrow ZZ \rightarrow e^+e^-e^+e^-\), will be recorded together with \(\sim 25\) other proton–proton interactions that occurred in the same bunch crossing. The interactions that make up the “underlying event” are often called “minimum-bias” interactions because they are the ones that would be selected by a trigger that selects interactions.
in an unbiased way. The presence of additional interactions that are recorded together with the one of interest is sometimes referred to as “pile-up”.

A further complication is that particle detectors do not have an infinitely fast response time – this is analogous to the exposure time of a camera. If the “exposure time” is shorter than the bunch-crossing period, the event will contain only information from the selected bunch crossing. Otherwise, the event will contain activity, if there is any, from neighbouring bunches in addition. In e+e− colliders (e.g. LEP) it is very unlikely for there to be any activity in nearby bunch crossings, which allows the use of slow detectors such as the time projection chamber (TPC). This is also true at HERA and in the ALICE experiment [1] at LHC that will study heavy-ion collisions at much lower luminosities than in the proton–proton case.

The bunch-crossing period for proton–proton collisions at LHC will be 25 ns (i.e. 40 MHz rate) – at the design luminosity the interaction rate will be \(O(10^8)\) Hz and there will be an average of about 25 interactions per bunch crossing. Some detectors, for example the ATLAS silicon tracker, achieve an exposure time of less than 25 ns, but many do not. For example, pulses from the ATLAS liquid-argon calorimeter extend over many bunch crossings (BCs).

4. DESIGN OF A TRIGGER AND DATA-ACQUISITION SYSTEM FOR LEP

Let us now return to the discussion of designing a T/DAQ system, considering the case of experiments at LEP (ALEPH [2], DELPHI [3], L3 [4], OPAL [5]), and building on the model developed in Section 2.

4.1 Using the bunch crossing as a “pre-trigger”

If the time between bunch crossings is reasonably long, one can use the clock that signals when bunches of particles cross as the pre-trigger. The first-level trigger can then use the time between bunch crossings to make a decision, as shown in Fig. 6. For most crossings the trigger will reject the event by issuing a fast clear – in such cases no dead-time is introduced. Following an “accept” signal, dead-time will be introduced until the data have been read out. This is the basis of the model that was used at LEP, where the bunch-crossing interval of 22 µs (11 µs in eight-bunch mode) allowed comparatively complicated trigger processing (latency ~few µs). Note that there is no (first-level) trigger dead-time because the decision is made during the interval between bunch crossings where no interactions occur.

In the following, the readout model used at LEP is illustrated concentrating on the example of ALEPH [2]. Figure 7 shows the readout model, using the same kind of block diagram as presented in Section 2. The BC clock is used to start the TDCs and generate the gate for the ADCs, and a first-level (LVL1) trigger decision arrives in less than 5 µs so that the fast clear can be completed prior to the next bunch crossing. For events retained by LVL1, a more sophisticated second-level (LVL2) trigger decision is made after a total of about 50 µs. Events retained by LVL2 are read out to local buffer memory (within the readout controllers or “ROCs”), and then passed to a global buffer. There is a final level of selection (LVL3) before recording the data on permanent storage for offline analysis.

---

3 The author was not involved in any of the LEP experiments. In these lectures the example of ALEPH is used to illustrate how triggers and data-acquisition were implemented at LEP; some numbers from DELPHI are also presented. The T/DAQ systems in all of the LEP experiments were conceptually similar.
2. Signals

1. BC clock

(Digitizer) Register

4. Fast read out

5. Final read out

3. Trigger (or fast clear)

Readout dead-time:
Trigger rate $\times$ local readout time
No trigger dead-time!

Figure 6: Readout system using BC clock and fast clear

For readout systems of the type shown in Fig. 7, the total dead-time is given by the sum of two components – the trigger dead-time and the readout dead-time:

- The trigger dead-time is evaluated by counting the number of BCs that are lost following each LVL1 trigger, and calculating the product of the LVL1 trigger rate, the number of lost BCs and the BC period. Note that the effective LVL2 latency, given by the number of lost BCs and the BC period, is less than (or equal to) the true LVL2 latency.

- The readout dead-time is given by the product of the LVL2 trigger rate and the time taken to perform local readout into the ROCs. Strictly speaking, one should also express this dead-time in terms of the number of BCs lost after the LVL2 trigger. Note that, as long as the buffers in the ROCs and the global buffers do not fill up, no additional dead-time is introduced by the LVL3 trigger.

Let us now look quantitatively for the example of the DELPHI experiment for which the T/DAQ system was similar to that described above for ALEPH. Typical numbers for LEP-II are shown in Table 1 [3].

4.2 Data acquisition at LEP

Let us now continue our examination of the example of the ALEPH T/DAQ system. Following a LVL2 trigger, events were read out locally and in parallel within the many readout crates – once the data had been transferred within each crate to the ROC, further LVL1 and LVL2 triggers could be accepted. Subsequently, the data from the readout crates were collected by the main readout computer, “building” a complete event. As shown in Fig. 8, event building was performed in two stages: an event contained a number of sub-events, each of which was composed of several ROC data blocks. Once a complete event was in the main readout computer, the LVL3 trigger made a final selection before the data were recorded.
Table 1
Typical T/DAQ parameters for the DELPHI experiment at LEP-II

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVL1 rate</td>
<td>~ 500 – 1000 Hz (instrumental background)</td>
</tr>
<tr>
<td>LVL2 rate</td>
<td>6 – 8 Hz</td>
</tr>
<tr>
<td>LVL3 rate</td>
<td>4 – 6 Hz</td>
</tr>
<tr>
<td>LVL2 latency</td>
<td>38 µs (1 lost BC ⇒ 22 µs effective)</td>
</tr>
<tr>
<td>Local readout time</td>
<td>~ 2.5 ms</td>
</tr>
<tr>
<td>Readout dead-time</td>
<td>~7 Hz \times 2.5 \times 10^{-3} s = 1.8%</td>
</tr>
<tr>
<td>Trigger dead-time</td>
<td>~750 Hz \times 22 \times 10^{-6} s = 1.7%</td>
</tr>
<tr>
<td>Total dead-time</td>
<td>~3 – 4%</td>
</tr>
</tbody>
</table>

The DAQ system used a hierarchy of computers – the local ROCs in each crate; event builders (EBs) for sub-events; the main EB; the main readout computer. The ROCs performed some data processing (e.g. applying calibration to convert ADC values to energies) in addition to reading out the data from ADCs, TDCs, etc. (Zero suppression was already performed at the level of the digitizers where appropriate.) The first layer of EBs combined data read out from the ROCs of individual sub-detectors into sub-events; then the main EB combined the sub-events for the different sub-detectors. Finally, the main readout computer received full events from the main EB, performed the LVL3 trigger selection, and recorded selected events for subsequent analysis.

As indicated in Fig. 9, event building was bus based – each ROC collected data over a bus from the digitising electronics; each sub-detector EB collected data from several ROCs over a bus; the main EB collected data from the sub-detector EBs over a bus. As a consequence, the main EB and the main readout computer saw the full data rate prior to the final LVL3 selection. At LEP this was fine – with an event rate after LVL2 of a few Hz and an event size ~100 kBytes, the data rate was a few hundred kBytes/s, much less than the available bandwidth (e.g. ~40 MBytes/s maximum on VME bus).
Figure 8: ALEPH data-acquisition architecture

Figure 9: Event building in ALEPH
4.3 Triggers at LEP

The triggers at LEP aimed to select any $e^+e^-$ annihilation event with a visible final state, including events with little visible energy, plus some fraction of two-photon events, plus Bhabha scattering events. Furthermore, they aimed to select most events by multiple, independent signatures so as to maximise the trigger efficiency and to allow the measurement of the efficiency from the data. The probability for an event to pass trigger A or trigger B is $\sim 1 - \delta_A\delta_B$, where $\delta_A$ and $\delta_B$ are the individual trigger inefficiencies, which is very close to unity for small $\delta$. Starting from a sample of events selected with trigger A, the efficiency of trigger B can be estimated as the fraction of events passing trigger B in addition.

5. PHYSICS REQUIREMENTS – TWO EXAMPLES

In the following, the physics requirements on the T/DAQ systems at LEP and at LHC are examined. These are complementary cases – at LEP precision physics was the main emphasis – at LHC discovery physics will be the main issue. Precision physics at LEP needed accurate determination of the absolute cross-section (e.g. in the determination of the number of light-neutrino species). Discovery physics at the LHC will require sensitivity to a huge range of predicted processes with diverse signatures (with very low signal rates expected in some cases), aiming to be as sensitive as possible to new physics that has not been predicted (by using inclusive signatures). This has to be achieved in the presence of an enormous rate of Standard Model physics backgrounds (the rate of proton–proton collisions at LHC will be $O(10^3) \text{ Hz} - \sigma \sim 100 \text{ mb}, L \sim 10^{13} \text{ cm}^2\text{s}^{-1}$).

5.1 Physics requirements at LEP

Triggers at LEP aimed to identify all events coming from $e^+e^-$ annihilations with visible final states. At LEP-I, operating with $\sqrt{s} \sim m_Z$, this included $Z \to$ hadrons, $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ and $Z \to \tau^+\tau^-$; at LEP-II, operating above the WW threshold, this included WW, ZZ and single-boson events. Sensitivity was required even in cases where there was little visible energy, e.g. in the Standard Model for $e^+e^- \to Z\gamma$ with $Z \to \nu\nu$, and in new-particle searches such as $e^+e^- \to \chi\bar{\chi}$ for the case of small $\chi^+ - \chi^0$ mass difference that gives only low-energy visible particles ($\chi^0$ lightest super-symmetric particle). In addition, the triggers had to retain some fraction of two-photon collision events (used for QCD studies), and identify Bhabha scatters (needed for precise luminosity determination).

The triggers could retain events with any significant activity in the detector. Even when running at the $Z$ peak, the rate of $Z$ decays was only $O(1) \text{ Hz} - \text{physics rate was not an issue. The challenge was in maximising the efficiency (and acceptance) of the trigger, and making sure that the small inefficiencies were very well understood. The determination of absolute cross-section depends on knowing the integrated luminosity and the experimental efficiency to select the process in question (i.e. the efficiency to trigger on the specific physics process). Precise determination of the integrated luminosity required excellent understanding of the trigger efficiency for Bhabha-scattering events (luminosity determined from the rate of Bhabha scatters within a given angular range). A major achievement at LEP was to reach per-mil precision.

The trigger rates (events per second) and the DAQ rates (Bytes per second) at LEP were modest as discussed in Section 4.

5.2 Physics requirements at LHC

Triggers in the general-purpose proton–proton experiments (ATLAS [6] and CMS [7]) will have to retain as high as possible a fraction of the events of interest for the diverse physics programmes of these experiments. Higgs searches in and beyond the Standard Model will include looking for $H \to ZZ \to$ leptons, $H \to \gamma\gamma$, and also $H \to \tau^+\tau^-$ and $H \to b\bar{b}$. Super-symmetry (SUSY) searches will be performed with and without the assumption of R-parity conservation. One will search for other new physics using inclusive triggers that one hopes will be sensitive to unpredicted processes. In parallel with the searches
for new physics, the LHC experiments aim to do precision physics, such as measuring the W mass, especially in the early phases of LHC running.

In contrast to the experiments at LEP, the LHC trigger systems have a hard job to reduce the physics event rate to a manageable level for data recording and offline analysis. As discussed above, the design luminosity $L \sim 10^{24}$ cm$^{-2}$s$^{-1}$, together with $\sigma \sim 100$ mb, implies $O(10^3)$ Hz interaction rate. Even the rate of events containing leptonic decays of W and Z bosons is $O(10^6)$ Hz. Furthermore, the size of the events is very large, $O(1)$ MByte, reflecting the huge number of detector channels and the high particle multiplicity in each event. Recording and subsequently processing offline $O(100)$ Hz event rate per experiment with $O(1)$ MByte event size is considered feasible, but it implies major computing resources [8]. Hence, only a tiny fraction of proton–proton collisions can be selected – taking the order-of-magnitude numbers given above, the maximum fraction of interactions that can be selected is $O(10^{-7})$. Note that the general-purpose LHC experiments have to balance the needs of maximising physics coverage and reaching acceptable (i.e. affordable) recording rates.

The LHCb experiment [9], which is dedicated to studying B-physics, faces similar challenges to ATLAS and CMS. It will operate at a comparatively low luminosity ($\sim 2 \cdot 10^{32}$ cm$^{-2}$s$^{-1}$), giving an overall proton–proton interaction rate of $\sim 20$ MHz – chosen to maximise the rate of single-interaction bunch crossings. The event size will be comparatively small ($\sim 100$ kBytes) as a result of having fewer detector channels and of the lower occupancy of the detector (due to the lower luminosity and hence less pile-up). However, there will be a very high rate of beauty production in LHCb – taking $\sigma \sim 500$ μb, the $b\bar{b}$ production rate will be $\sim 100$ kHz – and the trigger must search for specific B-decay modes that are of interest for physics analysis, with the aim of recording an event rate of only $\sim 200$ Hz.

The heavy-ion experiment ALICE [1] is also very demanding, particularly from the DAQ point of view. The total interaction rate will be much smaller than in the proton–proton experiments – $L \sim 10^{27}$ cm$^{-2}$s$^{-1}$ $\Rightarrow$ rate $\sim 8000$ Hz for Pb–Pb collisions. However, the event size will be huge due to the high final-state multiplicity in Pb–Pb interactions at LHC energy. Up to $O(10^4)$ charged particles will be produced in the central region giving an event size of up to $\sim 40$ MBytes when the full detector is read out. The ALICE trigger will select “minimum-bias” and “central” events (rates scaled down to a total of about 40 Hz), and events with dileptons ($\sim 1$ kHz with only part of the detector read out). Even more than in the other LHC experiments, the volume of data to be stored and subsequently processed offline will be massive, with a data rate to storage of $\sim 1$ GBytes/s (considered to be about the maximum affordable rate).

6. SELECTION CRITERIA AND TRIGGER IMPLEMENTATIONS AT LEP

The details of the selection criteria and trigger implementations at LEP varied from experiment to experiment [2, 3, 4, 5]. Discussion of the example of ALEPH is continued with the aim of giving a reasonably in-depth view of one system. For triggering purposes, the detector was divided into segments with a total of 60 regions in $\theta$, $\phi$ ($\theta$ is polar angle, $\phi$ is azimuth). Within these segments, the following trigger objects were identified:

1. muon – requiring a track penetrating the hadron calorimeter and seen in the inner tracker;

2. charged electromagnetic (EM) energy – requiring an EM calorimeter cluster and a track in the inner tracker;

3. neutral EM energy – requiring an EM calorimeter cluster (with higher thresholds than in (2) to limit the rate to acceptable levels).

In addition to the above local triggers, there were total-energy triggers (applying thresholds on energies summed over large regions – the barrel or a full endcap), a back-to-back tracks trigger, and triggers for Bhabha scattering (luminosity monitor).
The LVL1 triggers were implemented using a combination of analogue and digital electronics. The calorimeter triggers were implemented using analogue electronics to sum signals before applying thresholds on the sums. The tracking trigger looked for patterns of hits in the inner-tracking chamber (ITC) consistent with a track with $p_T > 1$ GeV \(^4\) – at LVL2 the TPC was used instead. The final decision was made by combining digital information from calorimeter and tracking triggers, making local combinations within segments of the detector, and then making a global combination (logical OR of conditions).

7. TOWARDS THE LHC

In some experiments it is not practical to make a trigger in the time between bunch crossings because of the short BC period – the BC interval is 132 ns at Tevatron-II, 96 ns at HERA and 25 ns at LHC. In such cases the concept of “pipelined” readout has to be introduced (also pipelined LVL1 trigger processing). Furthermore, in experiments at high-luminosity hadron colliders the data rates after the LVL1 trigger selection are very high, and new ideas have to be introduced for the high-level triggers (HLTs) and DAQ – in particular, event building has to be based on data networks and switches rather than data buses.

7.1 Pipelined readout

In pipelined readout systems (see Fig. 10), the information from each bunch crossing, for each detector element, is retained during the latency of the LVL1 trigger (several $\mu$s). The information may be retained in several forms – analogue levels (held on capacitors); digital values (e.g. ADC results); binary values (i.e. hit / no hit).

![Figure 10: Pipelined readout](image)

Pipelined readout systems will be used in the LHC experiments (they are already being used in existing experiments at HERA [10, 11] and the Tevatron [12, 13], but the demands at LHC are even greater due to the short BC period). A typical LHC pipelined readout system is illustrated in Fig. 11, where the digitiser and pipeline are driven by the 40 MHz BC clock. A LVL1 trigger decision is made for each bunch crossing (i.e. every 25 ns), although the LVL1 latency is several microseconds – the LVL1 trigger must concurrently process many events (this is achieved by using pipelined trigger processing as discussed below).

The data for events that are selected by the LVL1 trigger are transferred into a “derandomizer” – a memory that can accept the high instantaneous input rate (i.e. one word per 25 ns) while being read

\(^4\) Here, $p_T$ is transverse momentum (measured w.r.t. the beam axis); similarly $E_T$ is transverse energy.
out at the much lower average data rate (determined by the LVL1 trigger rate rather than the BC rate). In principle no dead-time needs to be introduced in such a system. However, in practice, data are retained for a few BCs around the one that gave rise to the trigger, and a dead period of a few BCs is introduced to ensure that the same data do not have to be accessed for more than one trigger. Dead-time must also be introduced to avoid the derandomizers from overflowing, e.g. where, due to a statistical fluctuation, many LVL1 triggers arrive in quick succession. The dead-time from the first of these sources can be estimated as follows (numbers from ATLAS): taking a LVL1 trigger rate of 75 kHz and 4 dead BCs following each LVL1 trigger gives $75 \text{ kHz} \times 4 \times 25 \text{ ns} = 0.75\%$. The dead-time from the second source depends on the size of the derandomizer and the speed with which it can be emptied – in ATLAS the requirements are < 1% dead-time for a LVL1 rate of 75 kHz (< 6% for 100 kHz).

![Diagram](https://via.placeholder.com/150)

**Figure 11:** Pipelined readout with derandomizer at LHC

Some of the elements of the readout chain in the LHC experiments have to be mounted on the detectors (and hence are totally inaccessible during running of the machine and in an environment with high radiation levels). This is shown for the case of CMS in Fig. 12.

![Diagram](https://via.placeholder.com/150)

**Figure 12:** Location of readout components in CMS

There are a variety of options for the placement of digitisation in the readout chain, and the optimum choice depends on the characteristics of the detector in question. Digitisation may be performed on the detector at 40 MHz rate, prior to a digital pipeline (e.g. CMS calorimeter). Alternatively, it may be done on the detector after multiplexing signals from several analogue pipelines.
(e.g. ATLAS EM calorimeter) – here the digitisation rate can be lower, given by the LVL1 trigger rate multiplied by the number of signals to be digitised per trigger. Another alternative (e.g. CMS tracker) is to multiplex analogue signals from the pipelines over analogue links, and then to perform the digitisation off-detector.

### 7.2 Pipelined LVL1 trigger

As discussed above, the LVL1 trigger has to deliver a new decision every BC, although the trigger latency is much longer than the BC period; the LVL1 trigger must concurrently process many events. This can be achieved by “pipelining” the processing in custom trigger processors built using modern digital electronics. The key ingredients in this approach are to break the processing down into a series of steps, each of which can be performed within a single BC period, and to perform many operations in parallel by having separate processing logic for each calculation. Note that in such a system the latency of the LVL1 trigger is fixed – it is determined by the number of steps in the calculation, plus the time taken to move signals and data to, from and between the components of the trigger system (e.g. propagation delays on cables).

Pipelined trigger processing is illustrated in Fig. 13 – as will be seen later, this example corresponds to a (very small) part of the ATLAS LVL1 calorimeter trigger processor. The drawing on the left of the figure depicts the EM calorimeter as a grid of “towers” in \( \eta - \phi \) space (\( \eta \) is pseudorapidity, \( \phi \) is azimuth angle). The logic shown on the right determines if the energy deposited in a horizontal or vertical pair of towers in the region \([A, B, C]\) exceeds a threshold. In each 25 ns period, data from one layer of “latches” (memory registers) are processed through the next step in the processing “pipe”, and the results are captured in the next layer of latches. Note that in the real system, such logic has to be performed in parallel for \(~3500\) positions of the reference tower, since “A” could be at any position in the calorimeter. In practice, modern electronics is capable of doing more than a simple add or compare operation in 25 ns, so there is more logic between the latches than in this illustration.

The amount of data to be handled varies with depth in the processing pipeline, as indicated in Fig. 14. Initially the amount of data expands compared to the raw digitisation level since each datum typically participates in several operations – the input data need to be “fanned out” to several processing elements. Subsequently the amount of data decreases as one moves further down the processing tree. The final trigger decision can be represented by a single bit of information for each BC – yes or no (binary 1 or 0). Note that, in addition to the trigger decision, the LVL1 processors produce a lot of data for use in monitoring the system and to guide the higher-levels of selection.
Although they have not been discussed in these lectures due to time limitations, some fixed-target experiments have very challenging T/DAQ requirements. Some examples can be found in Refs. [14, 15].

Many input data

- Energies in calorimeter towers (e.g. ~7000 trigger towers in ATLAS)
- Pattern of hits in muon detectors (e.g. O(10^6) channels in ATLAS)

Fan-out
(e.g. each tower participates in many calculations)

Tree
(Data for monitoring)

1-bit output
(YES or NO)

(Information to guide next selection level)

Figure 14: LVL1 data flow

8. SELECTION CRITERIA AT LHC

Features that distinguish new physics from the bulk of the cross-section for Standard Model processes at hadron colliders are generally the presence of high-$p_T$ particles (or jets). For example, these may be the products of the decays of new heavy particles. In contrast, most of the particles produced in minimum-bias interactions are soft ($p_T \sim 1$ GeV or less). More specific signatures are the presence of high-$p_T$ leptons (e, $\mu$, $\tau$), photons and/or neutrinos. For example, these may be the products (directly or indirectly) of new heavy particles. Charged leptons, photons and neutrinos give a particularly clean signature (c.f. low-$p_T$ hadrons in minimum bias events), especially if they are “isolated” (i.e. not inside jets). The presence of heavy particles such as W and Z bosons can be another signature for new physics – e.g. they may be produced in Higgs decays. Leptonic W and Z decays give a very clean signature that can be used in the trigger. Of course it is interesting to study W and Z boson production per se, and such events can be very useful for detector studies (e.g. calibration of the EM calorimeters).

In view of the above, LVL1 triggers at hadron colliders search for the following signatures:

- High-$p_T$ muons – these can be identified as charged particles that penetrate beyond the calorimeters; a $p_T$ cut is needed to control the rate of muons from $\pi^+ \rightarrow \mu^+\nu$ and $K^+ \rightarrow \mu^+\nu$ decays in flight, as well as those from semi-muonic beauty and charm decays.
- High-$p_T$ photons – these can be identified as narrow clusters in the EM calorimeter; cuts are made on transverse energy ($E_T >$ threshold), and isolation and associated hadronic transverse energy ($E_T <$ threshold), to reduce the rate due to misidentified high-$p_T$ jets.
- High-$p_T$ electrons – identified in a similar way to photons, although some experiments require a matching track already at LVL1.
- High-$p_T$ taus – identified as narrow clusters in the calorimeters (EM+hadronic energy combined).
- High-$p_T$ jets – identified as wider clusters in the calorimeters (EM+hadronic energy combined); note that one needs to cut at very high $p_T$ to get acceptable rates given that jets are the dominant high-$p_T$ process.
- Large missing $E_T$ or scalar $E_T$.

Some experiments also search for tracks from displaced secondary vertices at an early stage in the trigger selection.
The trigger selection criteria are typically expressed as a list of conditions that should be satisfied – if any of the conditions is met, a trigger is generated (subject to dead-time requirements, etc.). In these notes, the list of conditions is referred to as the “trigger menu”, although the name varies from experiment to experiment. An illustrative example of a LVL1 trigger menu for high-luminosity running at LHC includes the following (numbers are given for the case of ATLAS):

- one or more muons with $p_T > 20$ GeV (rate ~ 11 kHz);
- two or more muons each with $p_T > 6$ GeV (rate ~ 1 kHz);
- one or more $\gamma$ with $E_T > 30$ GeV (rate ~ 22 kHz);
- two or more $e/\gamma$ each with $E_T > 20$ GeV (rate ~ 5 kHz);
- one or more jets with $E_T > 290$ GeV (rate ~ 200 Hz);
- four or more jets $E_T > 90$ GeV (rate ~ 200 Hz).

The above represents an extract from a LVL1 trigger menu, listing some of the most important trigger requirements – the full menu would include many items in addition (typically ~100 items in total). The additional items are expected to include:

- $\tau$ (or isolated single-hadron) candidates;
- combinations of objects of different types (e.g. muon and $e/\gamma$);
- pre-scaled$^5$ triggers with lower thresholds;
- triggers needed for technical studies and to aid understanding of the data from the main triggers (e.g. trigger on bunch crossings at random to collect an unbiased data sample).

9. LVL1 TRIGGER DESIGN FOR LHC

A number of design goals must be kept in mind for LVL1 triggers at LHC. It is essential to achieve a very large reduction in the physics rate, otherwise the HLT/DAQ system will be swamped and the dead-time will become unacceptable. In practice, the interaction rate, $O(10^9)$ Hz, must be reduced to less than 100 kHz in ATLAS and CMS. Complex algorithms are needed to reject the background while keeping the signal events.

Another important constraint is to achieve a short latency – information from all detector channels ($O(10^7 - 10^8)$ channels!) has to be held on the detector pending the LVL1 decision. The pipeline memories that do this are typically implemented in ASICs (application-specific integrated circuits), and memory size contributes to the cost. Typical LVL1 latency values are a few microseconds (e.g. less than 2.5 $\mu$s in ATLAS and less than 3.2 $\mu$s in CMS).

A third requirement is to have flexibility to react to changing conditions (e.g. wide range of luminosities) and – hopefully – to new physics! The algorithms must be programmable, at least at the level of parameters (thresholds, etc.).

9.1 Case study – ATLAS $e/\gamma$ trigger

The ATLAS $e/\gamma$ trigger algorithm can be used to illustrate the techniques used in LVL1 trigger systems at LHC. It is based on 4x4 “overlapping, sliding windows” of trigger towers as illustrated in Fig. 15.

---

5 Some triggers may be “pre-scaled” – this means that only every $N^{th}$ event satisfying the relevant criteria is recorded, where $N$ is a parameter called the pre-scale factor. This is useful for collecting samples of high-rate triggers without swamping the T/DAQ system.
Each trigger tower has a lateral extent of 0.1×0.1 in $\eta$, $\phi$ space, where $\eta$ is pseudorapidity and $\phi$ is azimuth. There are about 3500 such towers in each of the EM and hadronic calorimeters. Note that each tower participates in calculations for 16 windows. The algorithm requires a local maximum in the EM calorimeter to define the $\eta$–$\phi$ position of the cluster and to avoid double counting of extended clusters (so-called “declustering”). It can also require that the cluster is isolated, i.e. little energy surrounding cluster in the EM calorimeter or in hadronic calorimeter.

The implementation of the ATLAS LVL1 calorimeter trigger is sketched in Fig. 16. Analogue electronics on the detector sums signals from individual calorimeter cells to form trigger-tower signals. After transmission to the “pre-processor” (PPr), which is located in an underground room, the tower signals are received and digitised; then the digital data are processed to obtain estimates of $E_T$ per trigger tower for each BC. At this point in the processing chain (i.e. at the output of the PPr), there is an “$\eta$–$\phi$ matrix” of the $E_T$ per tower in each of the EM and hadronic calorimeters that gets updated every 25 ns.

The tower data from the PPr are transmitted to the cluster processor (CP). Note that the CP is implemented with very dense electronics so that there are only four crates in total. This minimises the number of towers that need to be transmitted (“fanned out”) to more than one crate. Fan out is required for towers that contribute to windows for which the algorithmic processing is implemented in different crates. Also, within each CP crate, trigger-tower data need to be fanned out between electronic modules, and then between processing elements within each module. Considerations of connectivity and data-movement drive the design.

In parallel with the CP, a jet/energy processor (JEP) searches for jet candidates and calculates missing-$E_T$ and scalar-$E_T$ sums. This is not described further here.
A very important consideration in designing the LVL1 trigger is the need to identify uniquely the BC that produced the interaction of interest. This is not trivial, especially given that the calorimeter signals extend over many BCs. In order to assign observed energy deposits to a given BC, information has to be combined from a sequence of measurements. Figure 17 illustrates how this is done within the PPr (the logic is repeated ~7000 times so that this is done in parallel for all towers). The raw data for a given tower move along a pipeline that is clocked by the 40 MHz BC signal. The multipliers together with the adder tree implement a finite-impulse-response filter whose output is passed to a peak finder (a peak indicates that the energy was deposited in the BC presently being examined) and to a look-up table that converts the peak amplitude to an $E_T$ value. Special care is taken to avoid BC misidentification for very large pulses that may get distorted in the analogue electronics, since such signals could correspond to the most interesting events. The functionality shown in the figure is implemented in ASICs (four channels per ASIC).

![Level-1 Calorimeter Trigger Architecture](image)

Figure 16: Overview of the ATLAS LVL1 calorimeter trigger
The transmission of the data (i.e. the $E_t$ matrices) from the PPr to the CP is performed using a total of ~5000 digital links each operating at 400 Mbits/s (each link carries data from two towers using a technique called BC multiplexing [6]). Where fan out is required, the corresponding links are duplicated with the data being sent to two different CP crates. Within each CP crate, data are shared between neighbouring modules over a very high density crate back-plane (~800 pins per slot in a 9U crate; data rate of 160 Mbits/s per signal pin using point-to-point connections). On each of the modules, data are passed to eight large field programmable gate arrays (FPGAs) that perform the algorithmic processing, fanning out signals to more than one FPGA where required.

As an exercise, it is suggested that students make an order-of-magnitude estimate of the total bandwidth between the PPr and the CP, considering what this corresponds to in terms of an equivalent number of simultaneous telephone calls.

The $e/\gamma$ (together with the $\gamma/h$) algorithms are implemented using FPGAs. This has only become feasible thanks to recent advances in FPGA technology since very large and very fast devices are needed. Each FPGA handles an area of 4x2 windows, requiring data from 7x5 towers in each of the EM and hadronic calorimeters. The algorithm is described in a programming language (e.g. VHDL) that can be converted into the FPGA configuration file. This gives flexibility to adapt algorithms in the light of experience – the FPGAs can be reconfigured in situ. Note that parameters of the algorithms can be changed easily and quickly, e.g. as the luminosity falls during the course of a fill of the LHC machine, since they are held in registers inside the FPGAs that can be modified at run time (i.e. there is no need to change the “program” in the FPGA).

10. HIGH-LEVEL TRIGGERS AND DAQ AT LHC

In the LHC experiments, data are transferred after a LVL1 trigger accept decision to large buffer memories – in normal operation, the subsequent stages should not introduce further dead-time. At this point in the readout chain, the data rates are still massive. An event size of ~1 MByte (after zero suppression / data compression) at ~100 kHz event rate gives a total bandwidth of ~100 GBytes/s. This
is far beyond the capacity of the bus-based event building of LEP. Such high data rates will be dealt with by using network-based event building and by only moving a subset of the data.

Network-based event building is illustrated in Fig. 18 for the example of CMS. Data are stored in the readout systems until they have been transferred to the filter systems (associated with high-level trigger (HLT) processing), or until the event is rejected. Note that no node in the system sees the full data rate – each readout system covers only a part of the detector – each filter system deals with only a fraction of the events.

![Figure 18: CMS event builder](image)

The LVL2 trigger decision can be made without accessing all of the data. Substantial rejection can be made with respect to LVL1 without accessing the inner tracking detectors – calorimeter triggers can be refined using the full-precision, full-granularity calorimeter information; muon triggers can be refined using the high-precision readout from the drift chambers of the muon spectrometer. It is therefore only necessary to access the inner tracking data for the subset of events that pass this initial selection. CMS and ATLAS both use this sequential selection strategy.

Another way to reduce the amount of data to be moved is the region-of-interest concept. Here, the LVL1 trigger indicates the geographical location in the detector of candidate objects. LVL2 then only needs to access data from the regions of interest, a small fraction of the total, even for the calorimeter and muon detectors that participated in the LVL1 selection.

There are trade-offs in designing the HLT/DAQ system, that are reflected in different choices made by different experiments. For example, the region-of-interest mechanism is used in ATLAS to access the data from the readout systems selectively, but not in CMS. This reduces by a substantial factor the amount of data that needs to be moved from the readout systems to the processors, but it implies relatively complicated mechanisms to serve the data selectively to the LVL2 trigger processors (requiring more complex software than in the CMS case).

Concerning hardware implementation, the computer industry is putting on the market technologies that can be used to build much of the HLT/DAQ systems at LHC. Computer network products now offer high performance at affordable cost. Personal computers (PCs) provide exceptional value for money in processing power, with high-speed network interfaces as standard items. Nevertheless, custom hardware is likely to be needed in the parts of the system that see the full LVL1 trigger output rate (~100 kHz). This concerns the readout systems that receive the detector data following a positive LVL1 trigger decision, and (in ATLAS) the interface to the LVL1 trigger that receives the region-of-interest pointers. Of course, this is in addition to the specialised front-end electronics associated with the detectors that was discussed earlier (digitisation, pipelines, derandomizers, etc.). As for the LVL1 trigger, the HLT has a trigger menu that describes which events
should be selected. This is illustrated in Table 2 for the example of ATLAS at high luminosity. It is worth noting that the total rate, 475 Hz, contains about 200 Hz of “physics” events (e.g. triggers due to real electrons and muons from W and Z decays that pass the selection criteria). Further work is going on to improve the algorithms and enhance the purity of the selection. However, there will certainly be a need to balance physics coverage (e.g. using inclusive triggers with reasonably low thresholds) against the enormous financial cost of the offline computing needed to store, process, transport and analyse the data.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimated high-level trigger rates (ATLAS)</strong></td>
</tr>
<tr>
<td>One or more electrons with ( p_T &gt; 30 \text{ GeV} ), or two or more electrons with ( p_T &gt; 20 \text{ GeV} )</td>
</tr>
<tr>
<td>One or more photons with ( p_T &gt; 60 \text{ GeV} ), or two or more photons with ( p_T &gt; 20 \text{ GeV} )</td>
</tr>
<tr>
<td>One or more muons with ( p_T &gt; 20 \text{ GeV} ), or two or more muons with ( p_T &gt; 10 \text{ GeV} )</td>
</tr>
<tr>
<td>One or more jets with ( p_T &gt; 580 \text{ GeV} ), or three or more jets with ( p_T &gt; 260 \text{ GeV} ), or four or more jets with ( p_T &gt; 150 \text{ GeV} )</td>
</tr>
<tr>
<td>One or more jets with ( p_T &gt; 100 \text{ GeV and } E_T &gt; 100 \text{ GeV} )</td>
</tr>
</tbody>
</table>

A major challenge lies in the HLT/DAQ software. The algorithms for the HLT can be subdivided, at least logically, into LVL2 and LVL3 trigger stages. These might be performed by two separate processor systems (e.g. ATLAS), or in two distinct processing steps within the same processor system (e.g. CMS). The algorithms have to be supported by a software framework that manages the flow of data, supervising an event from when it arrives at the HLT/DAQ system until it is either rejected, or it is accepted and recorded on permanent storage. This includes software for efficient transfer of data to the algorithms. In addition to the above, there is a large amount of associated online software (run control, databases, book-keeping, etc.).

11. CONCLUDING REMARKS

It is hoped that these lectures have succeeded in giving some insight into the challenges of building T/DAQ systems for HEP experiments. These include challenges connected with the physics (inventing algorithms that are fast, efficient for the physics of interest, and that give a large reduction in rate), and challenges in electronics and computing. It is also hoped that the lectures have demonstrated how the subject has evolved to meet the increasing demands, e.g. of LHC compared to LEP, using new ideas based on new technologies.

ACKNOWLEDGEMENTS

The author would like to thank the local organising committee for their hospitality in Brazil. He would also like to thank the following people for their help and advice in preparing the lectures and the present notes: Bob Blair, Helfried Burckhart, Vincenzo Canale, Philippe Charpentier, Eric Eisenhandler, Markus Elsing, Philippe Farthouat, John Harvey, Jim Linnerman, Claudio Luci, Wesley Smith.
REFERENCES

J. Schukraft, “Heavy Ions”, in these proceedings.


INTERNATIONAL SCIENTIFIC COMMITTEE

Alvaro De Rújula (CERN, Switzerland)
Oscar Eboli (IFT, UNESP, Brazil)
Claire Earnshaw (CERN, Switzerland), School Administrator
Nick Ellis (CERN, Switzerland)
Carlos Garcia Canal (UNLP, Argentina)
Egil Lillestøl (CERN, Switzerland), Schools Director
Bernard Marechal (UFRJ, Brazil), Local Director
Luis Masperi, Director of CLAF
Arnulfo Zepeda (CINVESTAV, Mexico)

LOCAL ORGANIZING COMMITTEE

Bernard Marechal (UFRJ), Chair/Local Director
Marcelo Chiapparini (UERJ)
Jussara M. De Miranda (CBPF)
Leandro De Paula (UFRJ)
Joao Des Anjos (CBPF)
Erasmo Ferreira (UFRJ)
Renata Zukanovich Funchal (USP)
Miriam Gandelman (UFRJ)
Marcelo Guzzo (Unicamp)
Rogerio Rosenfeld (UNESP)

LECTURERS

Andy Cohen (Boston University, USA)
Manuel Delfino Reznicek (CERN, Switzerland)
Nick Ellis (CERN, Switzerland)
Carlos Escobar (Unicamp, Brazil)
Belen Gavela Legazapi (Univ. Autonoma de Madrid, Spain)
Ronald Kleiss (University of Nijmegen, The Netherlands)
Paolo Lipari (INFN, Italy)
Michelangelo Mangano (CERN, Switzerland)
Estaban Roulet (CONICET, Argentina)
Jurgen Schukraft (CERN, Switzerland)
Olav Ullaland (CERN, Switzerland)

DISCUSSION LEADERS

Gilvan Alves (CBPF, Brazil)
Guillermo Contreras (CINVESTAV Merida, Mexico)
Marco Aurelio Diaz (PUC, Chile)
Maria Teresa Dova (UNLP, Argentina)
Roberto Martinez (Univ. Nacional, Bogota)
Miguel Angel Perez (CINVESTAV, Mexico)
Rodolfo Sassot (UBA, Argentina)
STUDENTS

Luis Fernando Alday
Diego Aristizabal Sierra
Juan Carlos Arteaga-Velazquez
Patricia Astorga
Henrique de Melo Jorge Barbosa
Mateo Alejandro Barkovich
Alan Barr
Marcos Betemps
Wahid Bhimji
Stewart Boogert
Patrick John Brockill
Eyal Brodet
Julian Candia
Alexandra Stephanie Careno
Fernando Catalani
Helder Chavez
Luz Adriana Cordero Cid
Eleazar Cuautle Flores
Roberto da Silva
Alejandro Daleo
Ana Helena De Campos
Jorgen D’Hondt
Rodolfo Alexander Diaz Sanchez
Armando Gallegos
Carlos A. Garcia
Enrique Garcia Reyes
Javier Gonzalez
Sjors Griipink
Claire Gwenlan
Hugo R. Hernandez-Mora
Jaime Hernandez-Sanchez
Simon Fernando Herrera Herrera
Gavin Hesketh
Ana Lima
Rene Luna
Ana Amelia Machado
Luiz Fernando Mackedanz
Analisa Gabriela Mariazzi
Jesus Martinez-Castro
Martin Medina-Gaytan
Alexandre Mesquita
Eduardo Moreno
Celio Adrega Moura (de Moura Junior)
Edivaldo Moura Santos
Edward Moysie
Gabriela Alejandra Navarro
Silvia Paganini
Jonathan Paley
Andres Pavololotzky
Ana das Merces Pelinson
Richard Penny
Andreas Peters
Erica Polycarpo (Ribeiro Polycarpo Macedo)
Juan Ponciano
Moises Razeira
Marcos Rigol
Veronica Riquer
J.-Alexis Rodriguez
Simon Rodriguez Rodriguez
Sergio Roman Lopez
Federico Andres Sanchez
Helena de Fatima Santos (Nunes Casimiro dos Santos)
Sven Schagen
Carlos Javier Solano Salinas
Luiz Vitor Souza Filho
Alejandro Andres Szynkman
Cecilia Uribe Estrada
Cristian Villavicencio Reyes
Roberval Walsh
Carlos Yaguna
Juan Francisco Zanella Beguelin
POSTERS

Authors

Alday, L. F.; Aldazabal, G.
ArteagaVelazquez, J.C.; VasquezLopez, C.; Zepeda, A.
Astorga, P.
Barbosa, H.M.J.
Barr, A.
Betemps, M.A.; Gay Ducati, M.B.; Ayala, A.L.
Bhimji, W.
Boogert, S.
Brodet, E.
Candia, J.; Epele, L.N.; Roulet, E.
Catalani, F.
D’Hondt, J.
Daleo, A.; Sassot, R.; Garcia Canal, C.
De Campos, A.H.
de Moura Junior, C.A; Moraes Guzzo, M.
Diaz, R.A.; Martinez, R.; Rodriguez, J.A.
Diaz, R.A.; Martinez, R.; Rodriguez, J.A.
Epele, L.; Gomez Dumm, D.; Szynkman, A.A.
Garcia, C.A.
Grijpink, S.
Gwenlan, C.
Harari, D.; Pavololotzky, A.; Zaldarriaga, M.
Hernandez Sanchez, J.; Diaz Cruz, J.L.; Toscano, S.J.
Hernandez, H.R.
Hesketh, G.

Poster title

Standard-like models from brane–antibrane systems
Diffuse reflectivity of Tyvek in air and water, and anisotropic effects
Deuteron photodisintegration
Fluorescence detectors for the Pierre Auger observatory
Baryon-violating, supersymmetry at the LHC
Nuclear effects in Drell–Yan process
Charmless hadronic $B$ decays to phi final states at BaBar
Strange-particle photoproduction at ZEUS
$A_{FB}$ measurement at LEP
Cosmic ray photo disintegration and the knee of the spectrum
Characterization of extensive air showers of high energy
$W$ mass at LEP
Factorization of semi-inclusive processes
Cosmological production of supermassive particles: a possible source of UHECR
Tau neutrinos in the Pierre Auger experiment
Bounds for lepton-flavour violation in the two-Higgs-doublet model using $g_\mu - 2$
The decay $t \to c\gamma$ in the two-Higgs-doublet model
CP violation beyond the Standard Model in $B$ meson decays
On the efficiency of an extensive-air-shower detector array
Charged-current deep-inelastic scattering
Measurement of multijet events in photoproduction at HERA
Non gaussianity in the CMB induced by radio point sources
An effective-Lagrangian description of charged Higgs decays $H^+ \to W^+\gamma$, $W^+ Z$ and $W^+ h^0$
Experiment 831 search for the rare decay $D^0 \to \mu^+ \mu^-$
An overview of the Fermilab & D0 upgrades, and some prospects for Run 2

281
Lima, A.
Lucio, J.L.; Gallegos, A.; Moreno, G.; Escribano, R.; Pesticou, J.
Machado, A.A.; Pisano, F.
Mariazzi, A.; Dova, M.T.
Martinez Castro, J.A.
Mesquita, A.

Moura Santos, E.
Moyse, E.
Navarro, G.
Paganini, S.
Paley, J.M.

Pelinson, A.; Pabris, J.C.; Shapiro, I.Z.
Peters, A.
Polycarpo, E.
Ponciano J.
Razeira, M.; Mesquita, A.; Vasconcellos, C.A.Z.; Pacheco, J. F.
Riquer, V.

Roman Lopez, S.; Fernandez Teller, A.; Gamez Flores, E.; Zepeda, A.
Santos, H.
Schagen, S.
Solano, J.

Souza Filho, L.V.

Uribe Estrada, C.
Walsh, R.

Quarks masses and mixings

Kaon photoproduction on deuterium
On the mass and width of the $f_0(980)$ and $\Phi(1020)$

Charged right-handed currents in the leptoquark bilepton flavor
Studies on inclined extensive air shower
Jet finding algorithm
A first study of the role of tidal forces in a compact stars binary
Quest for the dynamics of $\nu_\mu \rightarrow \nu_\tau$ conversion
ATLAS first-level calorimeter trigger simulation
Deuterium targets and parton densities
Muons observed by Eascamp central detector
A new measurement of the anomalous magnetic moment of the muon
On the gravitational waves on the background of anomaly-induced inflation
A new measurement of direct $CP$ violation
Study of the LHCb sensitivity to the decay $B_\circ \rightarrow \mu^+ \mu^-$
Padé summation of perturbation expansion
A quark–meson coupling model with a fuzzy bag: an application to the equation of state of neutron stars
Variational calculations on two-body relativistic problem
Test experimental set up for the CORAL proposal (preliminary results)
The NA50 experiment
Charm in the proton
Study of polarization of hyperons in 500 GeV $\pi^-$–nucleon interactions
Simulation of fluorescence light of extensive air shower and construction of the fluorescence telescope of the Pierre Auger observatory
Charm–anticharm photoproduction asymmetry for baryons at E831/FOCUS
Constraints on the anomalous $Z\gamma Z$ vertex at electron–electron colliders

282