Doctoral Thesis

Models in Neutrino Physics

Numerical and Statistical Studies

Johannes Bergström

Theoretical Particle Physics, Department of Theoretical Physics,
School of Engineering Sciences
Royal Institute of Technology, SE-106 91 Stockholm, Sweden

Stockholm, Sweden 2013

Typeset in LaTeX

Akademisk avhandling för avläggande av teknologie doktorsexamen (TeknD) inom ämnesområdet fysik.

Scientific thesis for the degree of Doctor of Philosophy (PhD) in the subject area of physics.

ISBN 978-91-7501-854-6

TRITA–FYS 2013:50
ISSN 0280-316X
ISRN KTH/FYS/--13:50–SE

Printed in Sweden by Universitetsservice US-AB, Stockholm, August 2013

Abstract

The standard model of particle physics can excellently describe the vast majority of data of particle physics experiments. However, in its simplest form, it cannot account for the fact that the neutrinos are massive particles and lepton flavors mixed, as required by the observation of neutrino oscillations. Hence, the standard model must be extended in order to account for these observations, opening up the possibility to explore new and interesting physical phenomena.

There are numerous models proposed to accommodate massive neutrinos. The simplest of these are able to describe the observations using only a small number of effective parameters. Furthermore, neutrinos are the only known existing particles which have the potential of being their own antiparticles, a possibility that is actively being investigated through experiments on neutrinoless double beta decay. In this thesis, we analyse these simple models using Bayesian inference and constraints from neutrino-related experiments, and we also investigate the potential of future experiments on neutrinoless double beta decay to probe other kinds of new physics.

In addition, more elaborate theoretical models of neutrino masses have been proposed, with the seesaw models being a particularly popular group of models in which new heavy particles generate neutrino masses. We study low-scale seesaw models, in particular the resulting energy-scale dependence of the neutrino parameters, which incorporate new particles with masses within the reach of current and future experiments, such as the LHC.

Keywords: Neutrino mass, lepton mixing, Majorana neutrinos, neutrino oscillations, neutrinoless double beta decay, statistical methods, Bayesian inference, model selection, effective field theory, Weinberg operator, seesaw models, inverse seesaw, right-handed neutrinos, renormalization group, threshold effects.

Sammanfattning

Standardmodellen för partikelfysik beskriver den stora majoriteten data från partikelfysikexperiment utmärkt. Den kan emellertid inte i sin enklaste form beskriva det faktum att neutriner är massiva partiklar och leptonsmakerna är blandande, vilket krävs enligt observationerna av neutrinooscillationer. Därför måste standardmodellen utökas för att ta hänsyn till detta, vilket öppnar upp möjligheten att utforska nya och intressanta fysikaliska fenomen.

Det finns många föreslagna modeller för massiva neutriner. De enklaste av dessa kan beskriva observationerna med endast ett fåtal effektiva parametrar. Dessutom är neutriner de enda kända befintliga partiklar som har potentialen att vara sina egna antipartiklar, en möjlighet som aktivt undersöks genom experiment på neutrinolöst dubbelt betasönderfall. I denna avhandling analyserar vi dessa enkla modeller med Bayesisk inferens och begränsningar från neutrinorelaterade experiment och undersöker även potentialen för framtida experiment på neutrinolöst dubbelt betasönderfall att bergänsa andra typer av ny fysik.

Även mer avancerade teoretiska modeller för neutrinomassor har föreslagits, med seesawmodeller som en särskilt populär grupp av modeller där nya tunga partiklar genererar neutrinomassor. Vi studerar seesawmodeller vid låga energier, i synnerhet neutrinoparametrarnas resulterande energiberoende, vilka inkluderar nya partiklar med massor inom räckhåll för nuvarande och framtida experiment såsom LHC.

Nyckelord: Neutrinomassor, leptonblandning, Majorananeutriner, neutrinooscillationer, neutrinolöst dubbelt betasönderfall, statistiska metoder, Bayesisk inferens, modellval, effektiv fältteori, Weinbergoperator, seesawmodeller, invers seesaw, högerhänta neutriner, renormeringsgrupp, tröskeleffekter.

Preface

This thesis is divided into two parts. Part I is an introduction to the subjects that form the basis for the scientific papers, while Part II consists of the five papers included in the thesis.

Part I of the thesis is organized as follows. In Chapter 1, a general introduction to the subject of particle physics is given. Chapter 2 deals with the standard model of particle physics and some simple extensions, with emphasis put on neutrino masses and lepton mixing. Chapter 3 introduces the experimental consequences of massive neutrinos, while Chapter 4 gives an overview of the seesaw models, treating in some detail the type I and inverse versions. Chapter 5 introduces the concepts of regularization and renormalization in quantum field theories and discusses renormalization group equations in seesaw models. Chapter 6 deals with statistical methods of data analysis, while Chapter 7 is a short summary of the results and conclusions found in the papers of Part II. Finally, in Appendix A, all the renormalization group equations of the type I seesaw model are given.

Note that Part II of the thesis should not be considered as merely an appendix, but as being part of the main text of the thesis. The papers include discussion and interpretation of the results presented in them. Since simple repetition of this material is deemed unnecessary, the reader is referred to the papers themselves for the results and the discussion, except for a short summary in Chapter 7. The background material presented in the first five chapters contains both a broader introduction of the considered topics, as well as a more detailed and technical description of the models and methods considered in the papers. Hence, although there is necessarily some overlap with the corresponding sections in the papers, the more detailed discussion should be of help to the reader unfamiliar with those topics.

List of papers included in this thesis

[1] J. Bergström, M. Malinský, T. Ohlsson, and H. Zhang
Renormalization group running of neutrino parameters in the inverse seesaw model
Physical Review **D81**, 116006 (2010)
`arXiv:1004.4628`

[2] J. Bergström, T. Ohlsson, and H. Zhang
Threshold effects on renormalization group running of neutrino parameters in the low-scale seesaw model
Physics Letters **B698**, 297 (2011)
arXiv:1009.2762

[3] J. Bergström, A. Merle, and T. Ohlsson
Constraining new physics with a positive or negative signal of neutrino-less double beta decay
Journal of High Energy Physics **05**, 122 (2011)
arXiv:1103.3015

[4] J. Bergström
Bayesian evidence for non-zero θ_{13} and CP-violation in neutrino oscillations
Journal of High Energy Physics **08**, 163 (2012)
arXiv:1205.4404

[5] J. Bergström
Combining and comparing neutrinoless double beta decay experiments using different nuclei
Journal of High Energy Physics **02**, 093 (2013)
arXiv:1212.4484

List of papers not included in this thesis

[6] J. Bergström and T. Ohlsson
Unparticle self-interactions at the Large Hadron Collider
Physical Review **D80**, 115014 (2009)
arXiv:0909.2213

Thesis author's contributions to the papers

Besides discussing methods, results, and conclusions of all the papers together with the other authors, the main contributions to the articles are

[1] I did a substantial part of the numerical computations, produced many of the plots, and did some of the analytical computations. I revised the manuscript and wrote some parts of it.

[2] I did all the analytical computations and wrote the corresponding sections of the manuscript. The contents of the manuscript and its revisions were decided upon together with the other authors.

[3] I did many of the numerical computations as well as the few analytical calculations which were involved. I wrote some parts of the manuscript and revised it.

[4] Single authored.

[5] Single authored.

Notation and Conventions

The metric tensor on Minkowski space that will be used is

$$(g_{\mu\nu}) = \mathrm{diag}(1, -1, -1, -1) \tag{1}$$

Dimensionful quantities will be expressed in units of $\hbar$ and c. Thus, one can effectively put $\hbar = c = 1$. As a result, both time and length are expressed in units of inverse mass,

$$[t] = T = M^{-1}, \quad [l] = L = M^{-1}.$$

Also, the Einstein summation convention is employed, meaning that repeated indices are summed over, unless otherwise stated.

Erratum

In paper I [1], there are factors of v^2, where v is the vacuum expectation value of the Higgs field, missing in Eq. (32). It should read

$$\begin{aligned} m_\nu\big|_{M_{i-1}} &\simeq bv^2 \left[\kappa + Y_\nu M_{\mathrm{R}}^{-1} M_S (M_{\mathrm{R}}^T)^{-1} Y_\nu^T\right]\big|_{M_i} + (a-b)v^2\kappa\big|_{M_i} \\ &= bm_\nu\big|_{M_i} + \Delta v^2 \kappa\big|_{M_i} \end{aligned}$$

Acknowledgments

I would like to thank my supervisor, Prof. Tommy Ohlsson, for giving me the opportunity to conduct research and do my PhD in particle phenomenology at KTH, and also for his advice, and the collaboration that resulted in our common papers included in this thesis. I would like to thank He Zhang, Alexander Merle, and Michal Malinský for interesting discussions and our collaboration during the beginning of my time at KTH.

To all my colleagues and friends at the Department of Theoretical Physics at KTH and the Department of Physics at Stockholm University: thanks for nice company and making AlbaNova a pleasant place to work in.

I want to thank my whole family and all my friends for reminding me of all the nice things in life which are not related to physics. Carl, Alex, Sara, and Filippa – I hope that with this thesis it is now clear that neutrinos could potentially be *Majorana* particles. Thanks to my Mamma and Pappa for their help and support, and Pappa also for borrowing me the summer house for going with colleagues and discussing physics in a relaxed environment. Mormor, Morfar and Sabine for all the nice weekends in the garden. Finally, I would like to thank my Natasha for her incredibly strong support and encouragement.

Johannes Bergström
Stockholm, August 2013

Contents

Part I

Introduction and background material

Chapter 1

Introduction

Physics, in its most general sense, is the study of the constituents of Nature and their properties, from the large size and age of the Universe to the very small distances and time scales associated with heavy elementary particles. The goal is to make experiments and observations, collect and organize the data, and construct theories or models to describe those data. This thesis is within the area of particle physics, which studies the smallest known building blocks of matter, the elementary particles.

A scientific theory must be able to make predictions which can be compared with experimental results, i.e., it must be possible to conduct experiments which could agree or disagree with the predictions of the theory. This should be possible not only in principle, but also in practice (at least in the not too distant future). In the end, a theory should be judged on how well it describes reality, and a good theory should not be inconsistent with the experimental data collected to date. However, it is not only whether a theory is inconsistent (or has been "falsified") or not which determines the validity of a theory. Observations which *confirm* the predictions of a model can increase its validity, but only in cases where it could have been falsified, but was not. A practical complication is that no experiment is perfect; there will always be uncertainties and noise which can often make it difficult to tell if an experiment actually confirmed a prediction, or contradicted it. This is especially common at the frontiers of physics, where one is looking for small signals not previously observed. This is the point where statistical analysis, taking into account these uncertainties and noise, is necessary in order to compare the predictions of a model with observations.

Occam's razor states in its most basic form states that, out of two models which can describe observations, the model which is the "simpler" one should be preferred over the more "complex" one. However, it is not at all clear what in general is meant by a "simple" model, although models which are extensions of a more basic model (to which the more complex reduces as a special case) are usually considered as more complex. However, we will see that when data is statistically

analyzed using what is called *Bayesian inference*, a quantitative form of Occam's razor emerges automatically from the laws of probability theory. More precisely, predictivity becomes the measure of simplicity, i.e., the model which best predicted the data is to be preferred. A simpler model making precise predictions before the data was observed is better than a model which is compatible with "anything" and thus actually predicted very little. In a sense, a more predictive model is more easily falsifiable, and so our confidence in it should increase if it is not falsified.

Particle physics is the study the most fundamental building blocks of the Universe, out of which all other objects are composed, and simple compositions of such building blocks. The elementary particles are the particles for which there exists no evidence of substructure. Thus, the property of being elementary is not really fixed, and particles once thought to be elementary could turn out not to be so in the future. Many of these particles are rather heavy, and so to produce and study them requires concentrating a lot of energy into a small region of space, and so the field also goes under the name of high-energy physics. These highly energetic particles can be created in man-made particle accelerators, but also in natural environments in the Universe and by us observed as cosmic rays.

A very good way to test theories of particle physics is to build machines, particle accelerators, that collide particles together and then observing what comes "flying" out in what directions and with what energies. Hence, in order to produce increasingly heavier particles, these accelerators need to be able to accelerate the particles to increasingly higher energies. The most powerful accelerator built so far is the Large Hadron Collider (LHC), which has been built in a circular tunnel 27 kilometers in circumference beneath the French-Swiss border near Geneva, Switzerland, and has been colliding particles for a few years. Its main goal is thus to look for new particles which we have not been able to find until now because the previous accelerators were not powerful enough.

Today, the established theory of the Universe on its most fundamental level is the *standard model* (SM) of particle physics. It describes all known fundamental particles and how they interact with each other, except for the gravitational interaction. It has been tested to great precision in a very large amount of experiments and has been found to be a good description of fundamental particles and their interactions at energies probed so far. Since its formulation in the 1960's, it has only been slightly modified. During its life, it has made a vast number of predictions which have later been confirmed by experiments. This includes the existence of new particles such as the Z- and W-bosons, the top quark, and the tau neutrino. Until last year, the only part of the standard model yet to be confirmed was the existence of the Higgs boson, which is related to the mechanism of generating the masses of the particles in the SM. However, after many years of intense efforts by the physics community, a new particle was discovered at the experiments ATLAS and CMS at the LHC. Since this particle seems to have just the properties which the Higgs is expected to have, it is probably the standard model Higgs boson so long searched for. More data could of course show signs of deviations from the standard model predictions, but none has been found so far. The LHC is also designed to search for

other new hypothetical particles. Many possible candidates for such particles have been suggested, with a very popular class of such particles being *supersymmetric* partners of the known SM particles. However, there are no signs of such new particles as of today. In addition, the LHC has made many measurements of processes predicted by the SM, and it has confirmed those prediction with high accuracy at previously unexplored energies. After a very successful initial run, the LHC is now being upgraded in order to be able to produce collisions with even higher energies.

Despite the incredible success of the SM, it has a number of shortcomings. First, gravity is not included in the SM, but is instead treated separately, usually using the general theory of relativity. Note that the SM is a *quantum* theory, while general relativity is inherently classical. Although it would be pleasant to have the SM and gravity unified in a full quantum theory, most such attempt lack testability, which is due to the fact that they only make unique predictions for processes at energies much higher than will ever be possible to study.

On a more practical level, cosmological and astrophysical observations indicate that there exist large amounts of massive particles in the Universe that have not been detected apart from their gravitational effects. Since none of the particles in the SM can constitute this *dark matter*, one expects that there are new particles waiting to be discovered in the future. However, since the nature of these particles is largely unknown, it is uncertain when (and if) they will be discovered.

Finally, there is the fact that experiments show that the particles known as *neutrinos* in the SM are massive. From the beginning the neutrinos were assumed to be massless in the SM, and hence some form of extension of the SM is now required. The fact that neutrinos are massive is the motivation for all the work presented in this thesis, which will be dedicated to the study of different extensions of the SM which can accommodate massive neutrinos and their experimental tests.

Chapter 2

The standard model of particle physics and slightly beyond

The standard model (SM) of particle physics is the currently accepted theoretical framework for the description of the elementary particles and their interactions. It has been tested to great precision in very many experiments and has been found to be a good description of fundamental particles and their interactions at the energies probed so far [7].

In this chapter an introduction to the SM is given. Emphasis is put on those aspects of the SM which are most relevant for the topics dealt with later in the thesis, i.e., the lepton sector in general and neutrino masses and lepton mixing in particular. First, the concept of a quantum field theory is introduced, followed by a review of the construction of the SM. Then, general fermion mass terms and the principles of effective quantum field theories are reviewed. Quark and lepton masses and mixing are treated and finally the discussion also goes slightly beyond the SM by including right-handed neutrinos. For reviews and deeper treatments of the SM, see, e.g., Refs. [8–12].

2.1 Quantum field theory

A classical field is a function associating some quantity to each point of space-time, and is an object with an infinite number of dynamical degrees of freedom. The SM is a *quantum field theory* (QFT), and as such it deals with the quantum mechanics of fields. Basically, this means that the classical fields are quantized, i.e., are promoted to operators. A classical field theory can be specified by a *Lagrangian*

density $\mathcal{L}$ (usually just called the Lagrangian), which is a function of the collection of fields $\Phi = \Phi(x)$, i.e., $\mathcal{L} = \mathcal{L}(\Phi(x))$. The *action* functional is given by

$$S[\Phi] = \int \mathcal{L}\, \mathrm{d}^D x \tag{2.1}$$

and gives the dynamics of the fields through the *Euler–Lagrange* equations of motion.

The QFTs we will study will basically also be defined by a Lagrangian. However, in QFT, one is not interested in the values of the fields themselves, which are not well-defined, but instead other quantities such as correlation functions and S-matrix elements. From these one can then calculate observable quantities such as cross-sections and decay rates of particles associated with the fields.

Symmetries and symmetry arguments have played and still play an important role in physics in general, and in QFT in particular. One important class of symmetries are *space-time symmetries*, which are symmetries involving the space-time coordinates. The QFTs we will consider will all be relativistic QFTs, meaning that the Lorentz group is a symmetry group of the theory. This implies that the fields we consider have to transform under some representation of the Lorentz group. The lowest dimensional representations correspond to the most commonly used types of fields,

- A *scalar field* has spin 0,
- A *spinor field* has spin 1/2,
- A *vector field* has spin 1.

Another kind of symmetries are *internal symmetries*, which are symmetries only involving the dynamical degrees of freedom, i.e., the fields, and not the space-time coordinates. A very important and useful class of such symmetries are the *gauge symmetries*, which will be the main principle behind the construction of the SM. Finally, the theories we consider will be *local*, which basically means that the Lagrangian is a local expression is the fields, i.e., that it only depends on the fields at a single space-time point.

The terms in the Lagrangian are usually classified as either

- A *kinetic term*, which is quadratic in a single field and involves derivatives,
- A *mass term*, which is quadratic in a single field and does not involve derivatives, or
- An *interaction term*, which involves more than two fields.

A constant term in the Lagrangian would essentially correspond to an energy density of the vacuum or a cosmological constant. Since this term is usually irrelevant for particle physics, it will not be discussed any further. Finally, there could also

be terms linear in a field (only for a scalar which is also a singlet under all other symmetries), implying that the minimum of the classical Hamiltonian is not at zero field value. Since these fields do not appear in the models we consider, neither this will be further mentioned.

2.2 Basic structure of the standard model

The SM is a gauge theory, and as such its form is dictated by the principle of gauge invariance. A gauge theory is defined by specifying the gauge group, the fermion and scalar particle content, and their representations. The gauge group for the SM is given by

$$G_{\mathrm{SM}} = \mathrm{SU}(3)_{\mathrm{C}} \otimes \mathrm{SU}(2)_{\mathrm{L}} \otimes \mathrm{U}(1)_{\mathrm{Y}}, \tag{2.2}$$

which is a twelve-dimensional *Lie group*. Here $\mathrm{SU}(3)_{\mathrm{C}}$ is the eight-dimensional gauge group of Quantum Chromodynamics (QCD), where the subscript stands for "color", which the corresponding quantum number is called. The group $\mathrm{SU}(2)_{\mathrm{L}} \otimes \mathrm{U}(1)_{\mathrm{Y}}$ is the four-dimensional gauge group of the Glashow–Weinberg–Salam model of weak interactions [13–15]. As will be described later, only the left-handed fermions are charged under the $\mathrm{SU}(2)_{\mathrm{L}}$ subgroup, and hence the subscript "L". The symbol "Y" represents the weak hypercharge. We will now proceed to describe the particles of the SM and their interactions.

2.2.1 The gauge bosons

The part of the SM Lagrangian containing the kinetic terms as well as the self-interactions of the gauge fields is determined by gauge invariance and is given by

$$\mathcal{L}_{\mathrm{gauge}} = -\frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{i,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.3}$$

where $a \in \{1, 2, \ldots, 8\}$, $i \in \{1, 2, 3\}$, and the field strength tensors are given in terms of the gauge fields as

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.4}$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g_2 \varepsilon^{ijk} W^j_\mu W^k_\nu, \tag{2.5}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_3 f^{abc} G^b_\mu G^c_\nu. \tag{2.6}$$

Here, (g_2, ε^{ijk}) and (g_3, f^{abc}) are the coupling and structure constants of $\mathrm{SU}(2)_{\mathrm{L}}$ and $\mathrm{SU}(3)_{\mathrm{C}}$, respectively. Note that gauge invariance excludes the possibility of a mass term for the gauge fields, and thus, the gauge bosons are massless. This is a problem, since some gauge bosons, i.e., the W-bosons and the Z-boson, are observed to be massive [7]. To incorporate massive gauge bosons, the gauge symmetry has to be broken in some way. This can, for example, be done through spontaneous symmetry breaking, in which case the fundamental Lagrangian, but not the vacuum, respects the symmetry.

2.2.2 The fermions

The next step in the construction of the SM is the introduction of the fermions and the specification of their charges. The fermions of the SM come in two groups, called *quarks* and *leptons*, which in turn come in three *generations* each. Given the representations of the fermion ψ, the kinetic term and the interactions with the gauge bosons are determined by the requirement of gauge invariance and are given by

$$\mathcal{L}_\psi = \mathrm{i}\overline{\psi}\not{D}\psi. \tag{2.7}$$

Here $\not{Q} = \gamma^\mu Q_\mu$, $\overline{\psi} = \psi^\dagger\gamma^0$, with γ^μ the Dirac gamma matrices, and

$$D_\mu = \partial_\mu - \mathrm{i}g_1 B_\mu Y - \mathrm{i}g_2 W_\mu^i \tau^i - \mathrm{i}g_3 G_\mu^a t^a \tag{2.8}$$

is the covariant derivative. The g_i's are the coupling constants corresponding to the different gauge groups, Y the hypercharge of ψ, the τ^i's the representation matrices under $\mathrm{SU}(2)_\mathrm{L}$, and the t^a's the representation matrices under $\mathrm{SU}(3)_\mathrm{C}$. Note that the hypercharge and representation matrices depend on which fermion is being considered, and that, if ψ is a singlet under some subgroup, then the generator of that group is zero when acting on ψ.

The fermion fields in the SM are all *chiral*, meaning that they transform under a specific representations of the Lorentz group. The two different kinds of chirality are *left-handed* or *right-handed*, as denoted by the subscripts "L" and "R". The *quark* fields are organized as

$$q_{\mathrm{L}i} = \begin{pmatrix} u_{\mathrm{L}i} \\ d_{\mathrm{L}i} \end{pmatrix},\ u_{\mathrm{R}i},\ d_{\mathrm{R}i},$$

where $i \in \{1, 2, 3\}$ is the generation index. They are all in the fundamental representation of $\mathrm{SU}(3)_\mathrm{C}$, while the left-handed $q_{\mathrm{L}i}$'s are doublets and the right-handed $u_{\mathrm{R}i}$'s and $d_{\mathrm{R}i}$'s are singlets of $\mathrm{SU}(2)_\mathrm{L}$. The *lepton* fields are all singlets of $\mathrm{SU}(3)_\mathrm{C}$ and organized as

$$\ell_{\mathrm{L}i} = \begin{pmatrix} \nu_{\mathrm{L}i} \\ e_{\mathrm{L}i} \end{pmatrix},\ e_{\mathrm{R}i},$$

where the $\ell_{\mathrm{L}i}$'s are doublets and the $e_{\mathrm{R}i}$'s are singlets of $\mathrm{SU}(2)_\mathrm{L}$. For both quarks and leptons, the names assigned to the components of the doublets correspond to the names of the fields which appear in the Lagrangian after the electroweak symmetry has been broken.

In order to restore the symmetry between the quark and lepton fields, one can also introduce the right-handed neutrinos $\nu_{\mathrm{R}i}$ in the list. However, they would be total singlets of the SM gauge group and are not needed to describe existing experimental data, and should thus be excluded in a minimal model.[1] The hypercharges

[1] The other right-handed fermions are seen directly in the interactions with the gauge bosons, since they are not gauge singlets, and they are also required for describing the masses of these fermions.

of all the fermions are given in such a way that the correct electric charges are assigned after the spontaneous breaking of the gauge symmetry.

The rest of this thesis will mostly be concerned with the electroweak sector of the SM, while QCD will not be discussed in detail. The electroweak interactions can largely be studied separately from QCD, since $SU(3)_{\mathrm{C}}$ remains unbroken and there is no mixing between the gauge fields of $\mathrm{SU(3)_C}$ and $SU(2)_{\mathrm{L}} \otimes \mathrm{U(1)_Y}$.

This concludes the introduction of the basic structure of the SM. However, in order to give a good description of experimental data, the gauge symmetry of the SM needs to be broken. Before the description of this breaking, a short summary of the different kinds of possible mass terms for fermions will be given.

2.3 Fermion mass terms

There are in general two types of mass terms for a fermion ψ that can be constructed, both giving the same kinematical masses. The first one is called a *Dirac mass term*, and has the form

$$-\mathcal{L}_{\mathrm{Dirac}} = m\overline{\psi}\psi. \tag{2.9}$$

However, the chiral fields included in the SM satisfy

$$\overline{\psi_{\mathrm{L/R}}}\,\psi_{\mathrm{L/R}} = 0 \tag{2.10}$$

due to the definition of chirality and $\overline{\psi}$, and thus terms on the form of Eq. (2.9) vanish for all the fields of the SM. One could try to remedy this by defining a new field

$$\chi \equiv \psi_{\mathrm{L}} + \psi_{\mathrm{R}}, \tag{2.11}$$

but in the SM, the left-handed and right-handed fields transform under different representations of $\mathrm{SU(2)_L}$, and thus the resulting mass term,

$$m\overline{\chi}\chi = m\left(\overline{\psi_{\mathrm{L}}}\psi_{\mathrm{R}} + \overline{\psi_{\mathrm{R}}}\psi_{\mathrm{L}}\right), \tag{2.12}$$

will not be gauge invariant.

To construct the second type of fermion mass term, called a *Majorana mass term*, one first would need to introduce the *charge conjugation operator* as

$$\hat{C} : \psi \to \psi^c = C\overline{\psi}^T, \tag{2.13}$$

where the matrix C satisfies

$$C^\dagger = C^T = C^{-1} = -C. \tag{2.14}$$

The Majorana mass term is then given by

$$-\mathcal{L}_{\mathrm{Majorana}} = \frac{1}{2}m\overline{\psi^c}\psi + \mathrm{H.c.}, \tag{2.15}$$

where "H.c." denotes the Hermitian conjugate, and m can always be made real and positive by redefining the phase of ψ. However, this kind of mass term is also

not gauge invariant, unless ψ is a gauge singlet. Any Abelian charges of ψ will be broken by two units. If the fermion ψ is chiral, as in the SM, the mass term can be rewritten as

$$-\mathcal{L}_{\text{Majorana}} = \frac{1}{2} m \overline{\xi} \xi, \tag{2.16}$$

where $\xi \equiv \psi + \psi^c$ is called a *Majorana field*, since it obeys $\xi^c = \xi$, called the *Majorana condition*. After field quantization, the Majorana condition on the field ξ will imply the equality of the particle and antiparticle states. A Majorana field has only half the independent components of a Dirac field.

In conclusion, none of the fermion fields in the *unbroken* SM can have a mass term, and thus, all SM fermions are massless. The only possible exception is the right-handed neutrino, which is a gauge singlet and can hence have a Majorana mass term. This is a problem, since the fermions existing in Nature are observed to be massive.[2]

2.4 The scalar sector and the Higgs mechanism

In order to make the model described above consistent with experiments, one needs to introduce some mechanism to break the SM gauge symmetry in a way that gives masses to the fermions and three of the gauge bosons. In the SM, this is achieved through the *Higgs mechanism* [16–21]. It is implemented by introducing one complex scalar $\text{SU}(2)_\text{L}$ doublet ϕ, called the *Higgs field*, which is described by the Lagrangian

$$\mathcal{L}_{\text{scalar}} = |D_\mu \phi|^2 - V(\phi), \tag{2.17}$$

where the *scalar potential* is given by

$$V(\phi) = -\mu^2 |\phi|^2 + \frac{\lambda}{4} |\phi|^4. \tag{2.18}$$

If $\mu^2 > 0$, the minimum of the potential will not be at $\phi = 0$, but instead where

$$|\phi| = v = \sqrt{\frac{2\mu^2}{\lambda}}, \tag{2.19}$$

which is called the *vacuum expectation value* (VEV) and experimentally determined to have a value of approximately 174 GeV.[3] This breaks electroweak gauge invariance and generates mass terms for the electroweak gauge bosons such that there

[2] The exceptions are the neutrinos, the masses of which have not been measured directly. However, the evidence for neutrino oscillations, to be discussed later, requires that they have small, but non-zero masses.

[3] Under standard conventions, the vacuum is such that $\langle \phi \rangle = (0 \; v)^T$.

are three massive gauge fields

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp \mathrm{i}W_\mu^2), \quad \text{with masses } m_W = g_2\frac{v}{\sqrt{2}}, \tag{2.20}$$

$$Z_\mu = \frac{1}{\sqrt{g_2^2+g_1^2}}(g_2 W_\mu^3 - g_1 B_\mu), \quad \text{with mass } m_Z = \sqrt{g_2^2+g_1^2}\frac{v}{\sqrt{2}}, \tag{2.21}$$

and one massless,

$$A_\mu = \frac{1}{\sqrt{g_2^2+g_1^2}}(g_1 W_\mu^3 + g_2 B_\mu). \tag{2.22}$$

The fields $W_\mu^\pm$, Z_μ, and A_μ are identified as the fields associated with the W-bosons, the Z-boson, and the photon, respectively.

The Higgs mechanism accomplishes the breaking

$$\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y} \to \mathrm{U}(1)_\mathrm{QED}, \tag{2.23}$$

where $\mathrm{U}(1)_\mathrm{QED}$ is the gauge group of Quantum Electrodynamics (QED). The fermion electric charge quantum number Q, i.e., the electric charge of a given fermion in units of the proton charge e, is given as

$$Q = T^3 + Y, \qquad e = \frac{g_1 g_2}{\sqrt{g_2^2+g_1^2}}, \tag{2.24}$$

where T^3 is the third component of the $\mathrm{SU}(2)_\mathrm{L}$ weak isospin. These assignments give the usual QED couplings of the fermions to the photon field, while the interactions with the W-bosons, i.e., the *charged-current* interactions, are given by[4]

$$\mathcal{L}_\mathrm{cc} = \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{u_\mathrm{L}}\gamma^\mu d_\mathrm{L} + \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{\nu_\mathrm{L}}\gamma^\mu e_\mathrm{L} + \mathrm{H.c.} \tag{2.25}$$

The introduction of a scalar field also opens up the possibility of further interactions with fermions through *Yukawa interactions*, having the form

$$-\mathcal{L}_\mathrm{Yuk} = \overline{\ell_\mathrm{L}}\phi Y_e e_\mathrm{R} + \overline{q_\mathrm{L}}\phi Y_d d_\mathrm{R} + \overline{q_\mathrm{L}}\tilde{\phi} Y_d d_\mathrm{R} + \mathrm{H.c.} \tag{2.26}$$

Here $\tilde{\phi} = \mathrm{i}\tau_2\phi^*$, where τ_2 is the second Pauli matrix, each fermion field is a vector consisting of the corresponding field from each generation, and Y_f for $f = e, u, d$ are Yukawa coupling matrices. When the Higgs field acquires its VEV, Dirac mass terms

$$-\mathcal{L}_\mathrm{mass} = \overline{e_\mathrm{L}} M_e e_\mathrm{R} + \overline{u_\mathrm{L}} M_u u_\mathrm{R} + \overline{d_\mathrm{L}} M_d d_\mathrm{R} + \mathrm{H.c.}, \tag{2.27}$$

are generated. Here the mass matrices

$$M_f = Y_f v \tag{2.28}$$

are arbitrary complex 3×3 matrices, and as such are in general not diagonal. In this case, the *flavor eigenstates*, which are the states participating in the weak

[4] There will also be interactions with the Z-boson, called neutral current interactions.

interactions, are not the same as the *mass eigenstates*, which are the states which propagate with definite masses. For n fermion generations, one would expect that each of the matrices Y_f contains n^2 complex , or $2n^2$ real, parameters. However, not all these parameters are physical, a point that will be discussed in more detail later.

One real degree of freedom of the Higgs fields is left as a physical field after the breaking of electroweak symmetry. The quantum of this field is usually called the *Higgs boson*. It could in principle have any mass, although an upper bound can be obtained by requiring the self-coupling constant to be perturbative. The Higgs boson has been actively searched for for many years, and one of the main reasons for the construction of the Large Hadron Collider at CERN was to study the mechanism of electroweak symmetry breaking, and, if that was the Higgs mechanism, find the Higgs boson. Then – finally – last year, a new particle was discovered at the experiments ATLAS [22] and CMS [23] at the LHC, with a mass of about 126 GeV. Since this particle seems to have the properties which the Higgs is expected to have in terms of its couplings, spin, etc., it is probably the SM Higgs boson. As always, more data could of course show signs of deviations from the SM predictions, but none has been found so far.

2.5 Effective field theory

So far, only terms in the Lagrangian with a small number of fields have been considered. From the kinetic term of a field, one can calculate its *mass dimension*. This is because, in natural units, the action in Eq. (2.1) is required to be dimensionless. Denote the mass dimension of X as $[X]$. If space-time is D-dimensional, then, since $[\,\mathrm{d}^D x] = -D$, the Lagrangian has to have mass dimension D. Using that $[\partial_\mu] = 1$, one obtains for the case $D = 4$

$$[\phi] = [A^\mu] = 1, \tag{2.29}$$

$$[\psi] = \frac{3}{2}. \tag{2.30}$$

From this, the mass dimensions of the constants multiplying all other terms in the Lagrangian can then be determined by the fact that the total mass dimension of each term must be 4.

It is now possible to further classify the interaction terms according to the mass dimension of the corresponding coupling constant. Field theory textbooks usually argue that a QFT should be "renormalizable", meaning that all divergences appearing should be possible to cancel with a finite number of counter terms. One can show that this is equivalent to having coupling constants with only non-negative mass dimensions, or equivalently, that the combinations of the fields in all terms in the Lagrangian have total mass dimension not greater than the space-time dimensionality. Otherwise, one needs an infinite number of counter terms, and hence, an

infinite number of unknown parameters, resulting in loss of predictive power of the theory.

An *effective* field theory Lagrangian, on the other hand, contains an infinite number of terms

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_D + \mathcal{L}_{D+D_1} + \mathcal{L}_{D+D_2} + \cdots , \tag{2.31}$$

where $\mathcal{L}_D$ is the renormalizable Lagrangian, $\mathcal{L}_{D+D_i}$ contains terms of dimension $D + D_i$, and $0 < D_1 < D_2 < \cdots$. Although there is an infinite number of terms in $\mathcal{L}_{\text{EFT}}$, one still has *approximate* predictive power. The coupling constants in $\mathcal{L}_{D+D'}$ have the form $g\Lambda^{-D'}$, where g is dimensionless and Λ is some energy scale. The amplitude resulting from this interaction at some energy scale $E < \Lambda$ will then be proportional to $g(E/\Lambda)^{D'}$. Thus, one can perform computations with an error of order $g(E/\Lambda)^{D''}$, with D'' the mass dimension of the next contributing term with higher dimension than D', if one keeps terms up to $\mathcal{L}_{D+D'}$ in $\mathcal{L}_{\text{EFT}}$. Hence, an effective field theory is just as useful as a renormalizable one, as long as one is satisfied with a certain finite accuracy of the predictions.[5] This also means that the leading contributions for a given process at low energies are induced by the operators of lowest dimensionality.

Given a renormalizable field theory involving a heavy field of mass M, one can integrate out the heavy field from the generating functional to produce an effective theory with an effective Lagrangian below M, consisting of a tower of effective operators. For example, in QED, one can integrate out the electron field to produce an effective Lagrangian, the Euler–Heisenberg Lagrangian

$$\mathcal{L}_{\text{EH}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{a}{m_e^4}\left(F^{\mu\nu}F_{\mu\nu}\right)^2 + \frac{b}{m_e^4}F^{\mu\nu}F_{\nu\sigma}F^{\sigma\rho}F_{\rho\mu} + \mathcal{O}\left(\frac{F^6}{m_e^8}\right), \tag{2.32}$$

where the dimensionless constants a and b can be found explicitly in terms of the electromagnetic coupling constant. However, even if one has no idea of what the high-energy theory is, one can still write down this unique Lagrangian (with unknown a and b, treated as free parameters) by simply imposing Lorentz, gauge, charge conjugation, and parity invariance. In other words, the only effect of the high-energy theory is to give explicit values of the coupling constants in the low-energy theory, which are then functions of the parameters of the high-energy theory.

Also, note that perturbative renormalization of effective operators can be performed in the same way as for those usually called "renormalizable", as long as one chooses the renormalization scheme wisely and works to a given order in E/Λ. In other words, to a given order in E/Λ, the effective theory contains only a finite number of operators, and working to a given accuracy, the effective theory behaves for all practical purposes like a renormalizable quantum field theory: only a finite number of counter terms are needed to reabsorb the divergences [24]. For deeper treatments of effective field theory, see Refs. [24–27].

In conclusion, the Lagrangian of the SM can be considered to contain terms of arbitrary dimensionality, of which the usual renormalizable SM Lagrangian is

[5] The accuracy is in practice almost always finite anyway if one is using perturbation theory.

the lowest order low-energy approximation. The allowed terms are given by the requirements of gauge and Lorentz invariance and any other assumed symmetries. There is only one allowed dimension-five operator (which is the lowest possible dimensionality of an effective operator) in the SM. This operator gives the lowest-order contribution to neutrino masses, and will be discussed in Section 2.7.1.

2.6 Quark masses and mixing

The mass terms for quarks in Eq. (2.27), i.e.,

$$-\mathcal{L}_{\text{q-mass}} = \overline{u_{\text{L}}} M_u u_{\text{R}} + \overline{d_{\text{L}}} M_d d_{\text{R}} + \text{H.c.}, \tag{2.33}$$

couple different quark flavors to each other, i.e., they mix the quark flavors. To find the mass eigenstate fields, i.e., the fields of which the excitations are propagating states, define a new basis of the quark fields by

$$u_{\text{L}} = U_{\text{L}} u'_{\text{L}}, \quad u_{\text{R}} = U_{\text{R}} u'_{\text{R}}, \quad d_{\text{L}} = V_{\text{L}} d'_{\text{L}}, \quad d_{\text{R}} = V_{\text{R}} d'_{\text{R}}, \tag{2.34}$$

where $U_{\text{L}}, U_{\text{R}}, V_{\text{L}},$, and V_{R} are some unitary 3×3 matrices. The kinetic terms of the quark fields are still diagonal in this new basis. Then, choose the unitary matrices such that

$$U_{\text{L}}^\dagger M_u U_{\text{R}} = D_u, \quad V_{\text{L}}^\dagger M_d V_{\text{R}} = D_d, \tag{2.35}$$

where D_u and D_d are real, positive, and diagonal. This choice of unitary matrices is possible for any complex matrices M_u and M_d. Thus, the fields

$$u'_i = u'_{\text{L}i} + u'_{\text{R}i}, \quad d'_i = d'_{\text{L}i} + d'_{\text{R}i} \tag{2.36}$$

are Dirac mass eigenstate fields with masses $m_{u,i} = (D_u)_{ii}$ and $m_{d,i} = (D_d)_{ii}$, respectively.

The interactions of the quarks with the gauge bosons originating from Eq. (2.25) will no longer be diagonal, but instead given by

$$\begin{aligned} \mathcal{L}_{Wud} &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{u_{\text{L}}} \gamma^\mu d_{\text{L}} + \text{H.c.} \\ &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{u'_{\text{L}}} U_{\text{L}}^\dagger V_{\text{L}} \gamma^\mu d'_{\text{L}} + \text{H.c.} \\ &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{u'_{\text{L}}} U_{\text{CKM}} \gamma^\mu d'_{\text{L}} + \text{H.c.}, \end{aligned} \tag{2.37}$$

where the unitary matrix $U_{\text{CKM}} = U_{\text{L}}^\dagger V_{\text{L}}$ is the *Cabibbo-Kobayashi-Maskawa* (CKM) or *quark mixing* matrix [28, 29].

A general complex $n \times n$ matrix can be parameterized by $2n^2$ real parameters, and a unitary one such as U_{CKM} by n^2 real parameters, out of which $n(n-1)/2$ are mixing angles and $n(n+1)/2$ are phases. By rephasing the left-handed quark fields,

one can naively remove $2n$ phases from the CKM matrix. However, since a global phase redefinition of the quark mass eigenstates leaves the CKM matrix invariant, only $(2n-1)$ phases can be removed. If one then rephases the right-handed quark fields in the same way, the Lagrangian will be left invariant. This means that these phases of the quark fields are not observable, and that neither are the removed phases from the CKM matrix. To conclude, the number of physical parameters for n quark generations are $2n$ masses, $n(n-1)/2$ angles and $(n-1)(n-2)/2$ phases. The total number of physical parameters of the quark sector is thus (n^2+1), which is to be compared to the naive expectation of $2n^2$ for each Yukawa matrix, i.e., $4n^2$ in total.

The CKM matrix can be parametrized in many ways, but a common parametrization is, for three generations, given by

$$\begin{aligned} U_{\text{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & C_{23} & S_{23} \\ 0 & -S_{23} & C_{23} \end{pmatrix} \begin{pmatrix} C_{13} & 0 & S_{13}e^{-\mathrm{i}\Delta} \\ 0 & 1 & 0 \\ -S_{13}e^{\mathrm{i}\Delta} & 0 & C_{13} \end{pmatrix} \begin{pmatrix} C_{12} & S_{12} & 0 \\ -S_{12} & C_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} C_{12}C_{13} & S_{12}C_{13} & S_{13}e^{-\mathrm{i}\Delta} \\ -S_{12}C_{23} - C_{12}S_{23}S_{13}e^{\mathrm{i}\Delta} & C_{12}C_{23} - S_{12}S_{23}S_{13}e^{\mathrm{i}\Delta} & S_{23}C_{13} \\ S_{12}S_{23} - C_{12}C_{23}S_{13}e^{\mathrm{i}\Delta} & -C_{12}S_{23} - S_{12}C_{23}S_{13}e^{\mathrm{i}\Delta} & C_{23}C_{13} \end{pmatrix}, \end{aligned} \tag{2.38}$$

where $C_{ij} = \cos\Theta_{ij}$ and $S_{ij} = \sin\Theta_{ij}$, Θ_{12}, Θ_{23}, and Θ_{13} are the quark mixing angles, and Δ is the CP-violating phase. The values of the quark mixing parameters have been inferred from experiments and the mixing angles have been found to be relatively small, while the phase factor $e^{-\mathrm{i}\Delta}$ is complex (and not real), i.e., there is CP-violation in the quark sector [7].

2.7 Lepton masses and mixing

The mass term for the charged leptons in Eq. (2.27), i.e.,

$$-\mathcal{L}_{\text{e-mass}} = \overline{e_L} M_e e_{\text{R}} + \text{H.c.}, \tag{2.39}$$

can be diagonalized in the same way as the down quark mass term by defining

$$e_{\text{L}} = V_{\text{L}} e'_{\text{L}}, \quad e_{\text{R}} = V_{\text{R}} e'_{\text{R}}, \tag{2.40}$$

where V_{L} and V_{R} are unitary matrices such that

$$V_{\text{L}}^{\dagger} M_e V_{\text{R}} = D_e, \tag{2.41}$$

where D_e is real, positive, and diagonal. Then,

$$e'_i = e'_{\text{L}i} + e'_{\text{R}i} \tag{2.42}$$

are Dirac fields with masses $m_{e,i} = (D_e)_{ii}$. Also note that V_L diagonalizes $M_e M_e^\dagger$ and V_R diagonalizes $M_e^\dagger M_e$, i.e.,

$$V_\mathrm{L}^\dagger \left(M_e M_e^\dagger\right) V_\mathrm{L} = V_\mathrm{R}^\dagger \left(M_e^\dagger M_e\right) V_\mathrm{R} = D_e^2, \tag{2.43}$$

and that similar relations hold for the quark mass matrices M_u and M_d. If neutrinos would be massless, i.e., have no mass terms, one could define the rotated neutrino fields by

$$\nu_\mathrm{L} = V_\mathrm{L} \nu'_\mathrm{L}, \tag{2.44}$$

in which case the charged current interaction in Eq. (2.25) would still be diagonal.

2.7.1 Neutrino masses without right-handed neutrinos

In the SM, the left-handed neutrinos do not obtain their masses through Yukawa interactions as the other fermions do. This is because there is no need for the introduction of right-handed neutrinos to describe experimental data, and hence, in the spirit of simplicity, there are no such fields in the SM to which the neutrinos could couple.

A simple and economical way to describe neutrino masses in the SM is based on an effective operator of dimension five (see Sec. 2.5), sometimes called the *Weinberg operator* [30]. It is given by

$$-\mathcal{L}_\nu^{d=5} = \frac{1}{2} \left(\overline{\ell_\mathrm{L}} \phi\right) \kappa \left(\phi^T \ell_\mathrm{L}^c\right) + \mathrm{H.c.}, \tag{2.45}$$

and is the only dimension-five operator allowed by the SM symmetries. Here, κ is a complex 3×3 matrix having the dimension of an inverse mass. However, using the anticommutativity of the fermion fields, one can show that only the symmetric part of κ is physically relevant, i.e., κ can always be chosen to be symmetric. Since the Higgs field acquires a VEV, the term in Eq. (2.45) will lead to a Majorana mass term for the light neutrinos,

$$-\mathcal{L}_\mathrm{Maj,L} = \frac{1}{2} \overline{\nu_\mathrm{L}} M_\mathrm{L} \nu_\mathrm{L}^c + \mathrm{H.c.}, \tag{2.46}$$

with $M_\mathrm{L} = v^2 \kappa$. Just as the mass matrices considered previously, M_L is in general not diagonal. One can redefine the neutrino fields as

$$\nu_\mathrm{L} = U_\mathrm{L} \nu'_\mathrm{L}, \tag{2.47}$$

where U_L is chosen such that

$$U_\mathrm{L}^\dagger M_\mathrm{L} U_\mathrm{L}^* = D_\mathrm{L}, \tag{2.48}$$

with D_L real, positive, and diagonal. This diagonalization is always possible for symmetric M_L. Equations (2.40) and (2.47) imply that the interaction in Eq. (2.25)

takes the form

$$\begin{aligned}\mathcal{L}_{W\nu e} &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{\nu_{\rm L}} \gamma^\mu e_{\rm L} + \text{H.c.} \\ &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{\nu'_{\rm L}} U_{\rm L}^\dagger V_{\rm L} \gamma^\mu e'_{\rm L} + \text{H.c.} \\ &= \frac{g_2}{\sqrt{2}} W_\mu^+ \overline{\nu'_{\rm L}} U^\dagger \gamma^\mu e'_{\rm L} + \text{H.c.},\end{aligned} \tag{2.49}$$

where $U = V_{\rm L}^\dagger U_{\rm L}$ is the *lepton mixing matrix*, also referred to as the *Pontecorvo-Maki-Nakagawa-Sakata* (PMNS) matrix [31–33]. The difference to the quark sector is that the Majorana mass term is not invariant under rephasings of the mass eigenstate fields. Thus, the phases of the Majorana neutrino fields are physical and cannot be removed from the lepton mixing matrix. It follows that there are $(n-1)$ additional physical phases for n generations. The lepton mixing matrix is usually parametrized as

$$\begin{aligned}U &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{diag}\left(e^{i\rho}, e^{i\sigma}, 1\right) \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-\mathrm{i}\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{\mathrm{i}\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{\mathrm{i}\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{\mathrm{i}\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{\mathrm{i}\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} e^{\mathrm{i}\rho} & & \\ & e^{\mathrm{i}\sigma} & \\ & & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13}e^{\mathrm{i}\rho} & s_{12}c_{13}e^{\mathrm{i}\sigma} & s_{13}e^{-\mathrm{i}\delta} \\ -s_{12}c_{23}e^{\mathrm{i}\rho} - c_{12}s_{23}s_{13}e^{\mathrm{i}(\delta+\rho)} & c_{12}c_{23}e^{\mathrm{i}\sigma} - s_{12}s_{23}s_{13}e^{\mathrm{i}(\delta+\sigma)} & s_{23}c_{13} \\ s_{12}s_{23}e^{\mathrm{i}\rho} - c_{12}c_{23}s_{13}e^{\mathrm{i}(\delta+\rho)} & -c_{12}s_{23}e^{\mathrm{i}\sigma} - s_{12}c_{23}s_{13}e^{\mathrm{i}(\delta+\rho)} & c_{23}c_{13} \end{pmatrix},\end{aligned} \tag{2.50}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, θ_{12}, θ_{23}, and θ_{13} are the lepton mixing angles, δ the CP-violating Dirac phase, and σ and ρ are CP-violating Majorana phases. Note however that there exists different parametrizations, differing in the convention for the CP-violating phases. Hence, in addition to the well-measured charged lepton masses, there are 9 parameters in the lepton sector of the SM with the Weinberg operator, separated as 3 neutrino masses, 3 mixing angles, and 3 CP-violating phases.

In conclusion, the SM can incorporate massive neutrinos, while also indicating that they should be light, a reflection of the fact that the first tree-level mass term has a dimension equal to five and not less. Whatever high-energy theory one can come up with in which the neutrinos are Majorana particles, it must always reduce to the SM with the Weinberg operator at low energies (unless it is forbidden by some symmetry of the high-energy theory). However, writing κ as

$$\kappa = \frac{\tilde{\kappa}}{\Lambda_\nu}, \tag{2.51}$$

with $\tilde{\kappa}$ dimensionless and Λ_ν some energy scale, we have that, since $v \simeq 174$ GeV,[6]

$$\Lambda_\nu \simeq \frac{v^2}{m_\nu}\tilde{\kappa} \simeq 3 \cdot 10^{13}\ \text{GeV}\left[\frac{\text{eV}}{m_\nu}\right]\tilde{\kappa}. \tag{2.52}$$

Since experiments indicate that the neutrino mass scale m_ν is of the order of 1 eV or smaller (see chapter 3), this will imply that the scale Λ_ν will be very high (unless $\tilde{\kappa}$ is very small), out of reach of any foreseeable experiments. The scale Λ_ν could be within the reach of future experiments, if either $\tilde{\kappa}$ is very small, which can be natural, or if for some reason the Weinberg operator is forbidden and neutrino masses are instead the result of even higher-dimensional operators [34]. Note that this operator necessarily introduces one more mass scale into the theory, above which the effective description ceases to be valid. Hence, one can say that at some high-energy scale, some kind of new physics must appear.

2.7.2 Neutrino masses with right-handed neutrinos

In order to restore the symmetry in the particle content of the SM, one can choose to extend it by adding 3 right-handed neutrinos $\nu_{\text{R}i}$[7], often also denoted by $N_{\text{R}i}$ or just N_i. Then, a new set of Yukawa couplings are allowed,

$$-\mathcal{L}_{\text{Yuk},\nu} = \overline{\ell_{\text{L}}}\tilde{\phi}Y_\nu\nu_{\text{R}} + \text{H.c.}, \tag{2.53}$$

which after electroweak symmetry breaking yields a Dirac-type mass terms as

$$-\mathcal{L}_{\text{Dirac},\nu} = \overline{\nu_{\text{L}}}M_{\text{D}}\nu_{\text{R}} + \text{H.c.}, \tag{2.54}$$

with $M_{\text{D}} = Y_\nu v$. However, since the right-handed neutrinos are total singlets under the SM gauge group, they can have Majorana masses on the form

$$-\mathcal{L}_{\text{Maj,R}} = \frac{1}{2}\overline{\nu_{\text{R}}^c}M_{\text{R}}\nu_{\text{R}} + \text{H.c.}, \tag{2.55}$$

with M_{R} symmetric. Without loss of generality, one can always perform a basis transformation on the right-handed neutrino fields and work in the basis in which M_{R} is real, positive, and diagonal, i.e., $M_{\text{R}} = \text{diag}(M_1, M_2, M_3)$. Thus, the full Lagrangian describing the masses in the neutrino sector is given by

$$-\mathcal{L}_{\nu\text{-mass}} = \frac{1}{2}\overline{\nu_{\text{L}}}M_{\text{L}}\nu_{\text{L}}^c + \overline{\nu_{\text{L}}}M_{\text{D}}\nu_{\text{R}} + \frac{1}{2}\overline{\nu_{\text{R}}^c}M_{\text{R}}\nu_{\text{R}} + \text{H.c.} = \frac{1}{2}\overline{\Psi}\mathcal{M}_\nu\Psi^c + \text{H.c.}, \tag{2.56}$$

where

$$\Psi = \begin{pmatrix} \nu_{\text{L}} \\ \nu_{\text{R}}^c \end{pmatrix}, \quad \mathcal{M}_\nu = \begin{pmatrix} M_{\text{L}} & M_{\text{D}} \\ M_{\text{D}}^T & M_{\text{R}} \end{pmatrix}. \tag{2.57}$$

Thus, it now has the form of a Majorana mass term for the field Ψ, with a symmetric 6×6 Majorana mass matrix $\mathcal{M}_\nu$. Diagonalization of this matrix leads in general

[6] Since we are dealing with matrices, the eigenvalues of which are the physical masses, the individual components of M_{L} could be much larger than the eigenvalues if there are large cancellations.

[7] Although there are models with different numbers of $\nu_{\text{R}i}$'s, we stick to 3 in this section for simplicity and symmetry reasons.

to 6 Majorana mass eigenstates, each of which is a linear superposition of the left- and right-handed neutrinos, and vice versa.

An often considered special case is $M_{\mathrm{L}} = M_{\mathrm{R}} = 0$, resulting in 6 Majorana fields which can be combined into 3 Dirac fields. In this case, one again has the freedom to rephase the neutrino fields, just as in the quark sector. Thus, the form of the lepton mixing matrix is given by Eq. (2.50), but without the last matrix containing the Majorana phases, in analogy with the CKM matrix in Eq. (2.38). However, for this to be the case, the M_{R} term has to be forbidden by some additional exact symmetry. A new gauge symmetry will not work, since after this symmetry is broken, the Majorana mass term for the right-handed neutrino will in general be generated. In this respect, the neutrino sector of the SM is fundamentally different from the charged lepton and quark sectors. Also, the neutrino Yukawa couplings have to be very small, of the order of 10^{-11}. Although this is technically natural, unless the right-handed neutrinos have some additional kind of interaction, the right-handed neutrinos will in practice be undetectable.

Another special case, which as been studied intensively in the literature, is the case $M_{\mathrm{R}} \gg M_{\mathrm{D}}$, which will be further discussed in Ch. 4. This is possible, since M_{R} is not related to the electroweak symmetry breaking, while M_{D} is determined by the Higgs VEV. On the other hand, a commonly used naturalness criterion states that a number can be naturally small if setting it to zero increases the symmetry of the Lagrangian. Since without a Majorana mass for the right-handed neutrino, the U(1) of total lepton number is a symmetry of the Lagrangian, a small M_{R} is also natural. To conclude, neutrinos can be Dirac particles, in which case there are additional degrees of freedom in the form of right-handed neutrinos. If instead neutrinos are Majorana particles there is some new physics at the scale Λ_ν. Hence, some kind of physics beyond the SM is required to describe non-zero neutrino masses.

Chapter 3

Experimental signatures of massive neutrinos

The observation of neutrino oscillations imply that neutrinos are massive particles, indicating the existence of new physics beyond the SM. In this chapter, we give a basic introduction to the different ways the effects of neutrino masses can be searched for experimentally. Neutrino oscillations are sensitive to the mixing parameters and mass-squared differences while beta decay, cosmology, and neutrinoless double beta decay are probes sensitive to the absolute values of neutrino masses.

3.1 Neutrino oscillations

In a charged-current interaction, the left-handed component of the mass eigenstate e_i' will be produced. Defining this as a charged lepton, and the neutrino produced in association with it as a flavor eigenstate, the lepton mixing matrix U relates the neutrino flavor eigenstates $|\nu_\alpha\rangle$ and the mass eigenstates $|\nu_i\rangle$ as

$$|\nu_\alpha\rangle = U_{\alpha i}^* |\nu_i\rangle. \tag{3.1}$$

This will lead to *neutrino oscillations*, in which a neutrino of flavor α, produced in a charged-current interaction, can, after propagating a certain distance, be detected as a neutrino of a generally different flavor β [35–38]. Since the time evolutions of the mass eigenstates are simply given by multiplication of the exponential $\exp(-\mathrm{i}E_i t)$, the amplitude for this transition is

$$\mathcal{A}\left(\nu_\alpha \to \nu_\beta\right) = \langle \nu_\beta | U_{\alpha i}^* e^{-\mathrm{i}E_i t} | \nu_i \rangle = \langle \nu_j | U_{\beta j} U_{\alpha i}^* e^{-\mathrm{i}E_i t} | \nu_i \rangle = U_{\beta i} \left(U^\dagger\right)_{i\alpha} e^{-\mathrm{i}E_i t}, \tag{3.2}$$

giving the transition probability as $P\left(\nu_\alpha \to \nu_\beta\right) = |\mathcal{A}\left(\nu_\alpha \to \nu_\beta\right)|^2$. As it turns out, neutrino oscillations are not sensitive to all the parameters in the neutrino sector,

which are 7 for Dirac neutrinos and 9 for Majorana neutrinos. The sensitivity is restricted to 2 independent mass-squared differences $\Delta m^2_{31} \equiv m^2_3 - m^2_1$, $\Delta m^2_{21} \equiv m^2_2 - m^2_1$, the 3 mixing angles, and the CP-violating Dirac phase δ. Neutrino oscillations have been verified by experiments on solar [39–42], atmospheric [43,44], and artificially produced neutrinos [45–49]. These measurements in combination with others have yielded rather strong constraints on the three mixing angles and the two mass squared differences, but with essentially no information on the CP-violating phase δ. Furthermore, since the sign of Δm^2_{31} is not known, neither is the ordering of the masses. The neutrino masses are said to have either normal or inverted ordering, depending on whether $m_1 < m_2 < m_3$ or $m_3 < m_1 < m_2$. Although neutrino oscillations are in principle sensitive to the mass ordering, there is as of today no firm evidence for one or the other.

While two of the mixing angles were long required to be non-zero, the third one, θ_{13}, was allowed by the data to vanish. This is related to the fact that the oscillations observed until recently correspond to the dominant effective two-flavor oscillation modes, driven by two mass-squared differences and two relatively large mixing angles. The purpose of many current and future experiments is mainly to explore sub-leading effects. Recently, several experiments have reported new interesting results related to two previously unobserved modes of neutrino oscillations, both of which rely on θ_{13} being non-zero. The first such mode is $\nu_\mu \to \nu_e$ oscillations, which has been investigated by looking at tee appearance of electron neutrinos in a beam consisting of mainly muon neutrinos in MINOS [50] and T2K [51,52]. Although the indication for a non-zero θ_{13} from these data are not really statistically significant, they are considered rather robust since these measurements are connected to the another mode of oscillation recently discovered, the disappearance of electron antineutrinos produced in the cores of nuclear reactors. Following an earlier indication by Double Chooz [53], the observation was first made by Daya Bay [54]. After that, further data has corroborated that measurement [55–58]. As a result, the uncertainty in the value of θ_{13} has become very small.

Following these successful measurements, there are two main goals for current and future neutrino oscillation experiments. First, one would like to establish whether there is CP-violation in the lepton sector of the standard model and measure the value of the CP-violating Dirac phase. In fact, CP-violation in neutrino oscillations, which is a genuine three-flavor effect, is only possible for a non-zero value of θ_{13}. Then, there is the determination of the neutrino mass ordering, and also any realistic possibility to determining that relies on θ_{13} not being too small [59]. Therefore, the recent experimental results on θ_{13} will be of crucial importance for the feasibility and planning of future experiments aiming to determine the neutrino mass ordering or search for leptonic CP-violation. Because individual neutrino oscillation experiments cannot determine all the oscillation parameters simultaneously, there exists a long history of global fits of oscillation data, in which all experimental results are combined. A recent such fit are has been performed in Ref. [60], with the resulting preferred ranges of the oscillation parameters (given as approximate confidence intervals) given in Tab. 3.1. There is a small, but insignifi-

cant, deviation of the best-fit value of θ_{23} away from maximal mixing ($\theta_{23} = \pi/4$). To test whether θ_{23} is maximal or not, and in the latter case whether it is smaller or larger than $\pi/4$, is also a goal for future experiments.

Most of the relevant results to date is consistent with the three-neutrino mixing scheme. However, there are some experimental results which does not seem to be, the most important ones being the old results of the LSND experiment [61, 62] and the more recent MiniBooNE experiment [63]. One common interpretation of these results is that they could be due to the effect of *sterile neutrinos*, which are eV-scale neutrinos without any weak interaction, mixing with the SM neutrinos. However, this interpretation is not really consistent with other experiments [64], and so the implications of these results are currently unclear.

To summarize, although neutrino oscillation experiments give a large amount of information on the neutrino sector of the standard model and the associated parameters, they are not directly sensitive to the absolute values of the neutrino masses, cannot distinguish between Dirac and Majorana neutrinos nor give information on the values of any possible Majorana phases. Luckily, there are other types of experiments which have the potential to answer the above questions.

Parameter	Preferred range (3σ)
$\theta_{12}/^\circ$	$31.1 - 35.9$
$\theta_{23}/^\circ$	$35.8 - 54.8$
$\theta_{13}/^\circ$	$7.2 - 10.0$
$\Delta m^2_{21}/(10^{-5}\ \mathrm{eV}^2)$	$7.00 - 8.09$
$\Delta m^2_{31}/(10^{-3}\ \mathrm{eV}^2)$ (NO)	$2.28 - 2.70$
$-\Delta m^2_{32}/(10^{-3}\ \mathrm{eV}^2)$ (IO)	$2.24 - 2.65$

Table 3.1. Currently preferred values for the neutrino oscillation parameters, adopted from Ref. [60]. The ranges are approximate "3σ" confidence intervals, NO (IO) stands for normal (inverted) neutrino mass ordering and $\Delta m^2_{32} = \Delta m^2_{31} - \Delta m^2_{21}$.

3.2 Beta decay and cosmology

Since the two mass square differences are measured accurately, it is sufficient to measure one of the individual masses to accurately infer all three of them. Of course, this could also be done by measuring other combinations of masses and mixing parameters, as long as the associated errors are small enough. There are a number of different types of experiments which are sensitive to the absolute values of the neutrino masses.

The most direct method is studying the energy spectra of electrons emitted in beta decays of certain isotopes. Beta decay is the decay of a nucleus accompanied by the emission of an electron or a positron. In *beta-minus decay*, an electron is emitted together with an electron antineutrino when a nucleus with mass number

A (the number of nucleons) and atomic number Z (the number of protons) decays according to

$$(A, Z) \to (A, Z+1) + e^- + \overline{\nu_e}, \tag{3.3}$$

which on the level of the nucleons is essentially

$$n \to p + e^- + \overline{\nu_e}. \tag{3.4}$$

In *beta-plus* decay, a positron is emitted together with an electron neutrino in the process

$$p \to n + e^+ + \nu_e, \tag{3.5}$$

which, because of energy conservation, can only occur inside a nucleus. The final state positron can also be exchanged for an initial state electron, in which case the process is called *electron capture*. All these processes can be accurately described through the exchange of SM W-bosons, or, since the relevant energies are much lower than the W-boson mass, by using the standard four-fermion interaction between the proton, neutron, electron, and neutrino fields. (Or between the quark, electron, and neutrino fields in the quark-level description.) Historically, the properties of beta decay, specifically the apparent non-conservation of energy and angular momenta, was what led Wolfgang Pauli to suggest that there was an undetected neutral particle being emitted together with the electron.

These spectra of the emitted electrons are sensitive to the effective kinematical electron neutrino mass m_β, given by

$$m_\beta^2 = \sum_{i=1}^{3} |U_{ei}|^2 \, m_i^2 = m_1^2 c_{12}^2 c_{13}^2 + m_2^2 s_{12}^2 c_{13}^2 + m_3^2 s_{13}^2. \tag{3.6}$$

Hence, such experiments cannot yield any information on any CP-violating phases, but since the mixing angles and the mass square differences are rather well determined, there is potential for measuring the absolute values of the neutrino masses. There is no evidence of the effect associated with a non-zero m_β^2, and m_β is constrained to be roughly below 2.5 eV [65, 66]. Near-future experiments such as MARE [67] and KATRIN [68, 69] are aiming to detect the effect of a non-zero m_β, or in any case to reduce the upper bound to the order of 0.2 eV.

The values of neutrino masses can also be probed by cosmological observations. The effective sum $\Sigma = m_1 + m_2 + m_3$ of neutrino masses can be inferred from measurements of the cosmic microwave background (CMB) radiation when combined with results from other observations, such as those of high-redshift galaxies, baryon acoustic oscillations, and type Ia supernovae [70, 71]. In addition, CMB observations are sensitive to additional light particles (relativistic degrees of freedom) in the early universe, such as sterile neutrinos. Although the *South Pole Telescope* (SPT) has reported an indication for a large value of Σ [72], neither the data from the *Atacama Cosmology Telescope* [73], nor the recent high-precision data from *Planck* [74] give any support for this indication. Σ is constrained to below between

about 0.2 eV and 1 eV, depending on which cosmological model is assumed and which sets of data is used. There is no sign of any relativistic degrees of freedom in addition to those 3 present in the standard model.

3.3 Neutrinoless double beta decay

Another well-investigated process related to neutrino masses is that of neutrinoless double beta decay. In nuclei where ordinary single beta decay is forbidden for kinematical reasons, *double beta decay* ($2\nu\beta\beta$) can be the dominant process. In this process, a nucleus with mass number A and atomic number Z decays through the emission of two electrons and two antineutrinos according to

$$(A, Z) \to (A, Z+2) + 2e^- + 2\overline{\nu_e}. \tag{3.7}$$

Double beta decay has been observed in around 10 nuclei, and the corresponding half-lives are very long, typically of the order of 10^{19} to 10^{21} years. This decay is essentially described as two "simultaneous" single beta decay processes, which is also the reason why the decay rates are so small.

If neutrinos are Majorana particles, it may be possible for the same nuclei to undergo double beta decay *without* emission of neutrinos. Replacing the two external neutrinos with an internal line and working on the level of the quarks inside the nucleons, one obtains the diagram in Fig. 3.1, giving the process

$$(A, Z) \to (A, Z+2) + 2e^-. \tag{3.8}$$

In this decay process, lepton number is violated by two units. This will be referred to as the "standard" mechanism responsible for $0\nu\beta\beta$ [75, 76]. Just as in the case of single beta decay, the internal momentum in the diagram is of the order of the typical energy transfer in the nucleus, and hence much smaller than the mass of the W-bosons. Thus, the quark-lepton interaction becomes point-like and can be described using the standard four-fermion interaction. However, due to the small neutrino masses, the light neutrino propagator will depend strongly on the energy transfer and can thus not be treated as point-like. In fact, due to the lightness of the neutrinos and the chirality structure of the charged current interaction vertices, the propagator of a Majorana mass eigenstate neutrino field with mass m_i will be [77]

$$P_{\mathrm{L}} \frac{\not{q} + m_i}{q^2 - m_i^2} P_{\mathrm{L}} = \frac{m_i}{q^2 - m_i^2} \simeq \frac{m_i}{q^2}, \tag{3.9}$$

where q is the transfered momentum.

In the calculation of the resulting inverse half-life T^{-1} of a specific nucleus N,[1] one can separate the dependence on the underlying particle physics and the nuclear physics by writing it as [75, 76]

$$T_N^{-1} = G_N |\mathcal{M}_N|^2 m_{ee}^2. \tag{3.10}$$

[1] The decay rate is in general equal to the inverse half-life divided by $\log 2$.

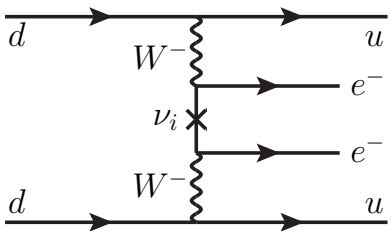

Figure 3.1. The leading order Feynman diagram for $0\nu\beta\beta$ through Majorana neutrino exchange.

Here, G_N is a known phase space factor, $\mathcal{M}_N$ is the *nuclear matrix element* (NME), containing all the dependence on the nuclear physics, and $|m_{ee}|$ is the effective neutrino mass given by

$$|m_{ee}| = \left|\sum_{i=1}^{3} U_{ei}^2 m_i\right| = \left|m_1 c_{12}^2 c_{13}^2 + m_2 s_{12}^2 c_{13}^2 e^{2i\alpha} + m_3 s_{13}^2 e^{2i\beta}\right|, \tag{3.11}$$

which is the magnitude of the ee-element of the neutrino mass matrix in the flavor basis. Here, α and β are the Majorana phases. This expression is given using a slightly different, but physically equivalent, parametrization of the lepton mixing matrix than what was used in Eq. (2.50). First, the neutrino fields have been given a common phase redefinition in order to make U_{e1} real. Then, the third neutrino mass eigenstate ν_3 has been given an additional phase redefinition so that the U_{e3} becomes independent of δ. From Eqs. 3.10 and 3.11 one observes that all the parameters which cannot be probed in oscillation experiments, i.e., the absolute values of the neutrino neutrino masses and the Majorana phases, could in principle be constrained using $0\nu\beta\beta$ experiments. The NME is given as the sum of two more basic matrix elements, the *Gamow–Teller* and *Fermi* type matrix elements as

$$\mathcal{M} = \mathcal{M}_{\mathrm{GT}} - \frac{g_{\mathrm{V}}^2}{g_{\mathrm{A}}^2}\mathcal{M}_{\mathrm{F}}, \tag{3.12}$$

where g_{V}^2 and g_{A}^2 are two constants of order one. The matrix elements $\mathcal{M}_{\mathrm{GT}}$ and $\mathcal{M}_{\mathrm{F}}$ can be written as expectation values of certain operators between the initial and final nuclear states [75, 76]. However, since they are rather complicated and not needed for the discussion of the particle physics, they will not be discussed in detail.

In order to extract the values of the underlying particle physics parameters, one needs the values of the NMEs. The calculation of the matrix elements $\mathcal{M}_{\mathrm{GT}}$ and $\mathcal{M}_{\mathrm{F}}$ requires the knowledge of the wave functions of complicated nuclei and need to be calculated numerically using some nuclear physics model. This is a notoriously difficult task [78–80], and the resulting *theoretical* uncertainties must be taken into account when inferring the parameters of the underlying models.

In addition, an observation of $0\nu\beta\beta$ will actually always imply that neutrinos are Majorana particles. This is because any diagram leading to $0\nu\beta\beta$, regardless of its origin or form, can be extended by connecting the external electron lines with the quark lines by two W-boson lines [81]. The resulting diagram will then be a four-loop diagram for a Majorana neutrino mass term. As of today, there is experimentally no other realistic way to determine if neutrinos are Dirac or Majorana particles.

In order to detect this extremely rare process against the much larger background of ordinary double beta decay, one can make use of the fact that the energy spectra of the final state electron differ widely. In $2\nu\beta\beta$, the final state contains four particles in addition to the nucleus, out of which only two (the electrons) can realistically be detected, while the neutrinos carry their energy out of the detector. Thus, the energy of the electrons will be broadly distributed in the region between 0 and the total energy being released, the *Q-value*. In contrast, due to the nucleus being very heavy compared to the electrons, most of the energy will be carried off by the electrons in $0\nu\beta\beta$, while being emitted almost back-to-back and monochromatically, and with the total energy equal to the Q-value. Using this, one can effectively discriminate the signal ($0\nu\beta\beta$) from the background ($2\nu\beta\beta$).

There has been no clear detection of the neutrinoless decay generally accepted by the community, although a small subset of the Heidelberg-Moscow collaboration, using ^{76}Ge as the decaying nucleus, did make such a claim [82]. This analysis was subsequently updated in Refs. [83, 84] (see also Ref. [85]). The claimed signal would correspond to m_{ee} being roughly equal to 0.3 eV. For many years, there was no other experiment with enough sensitivity to test these claims. Recently, however, the first data from a new generation of experiments has been released by EXO [86] and KamLAND-Zen [87], which have sensitivities in the same region of m_{ee}-values as those preferred by the ^{76}Ge claim. These experiments find no evidence of the decay and the reported combined lower limit on the half-life is $3.4 \cdot 10^{25}$ yr at 90% C.L., which corresponds to upper limits on m_{ee} in the range $0.12\text{ eV} - 0.25$ eV, depending on the NME assumed. It is not only the NMEs which are uncertain, however, but also the upper limit by itself might not accurately represent the constraints implied by the data. The reason is that the limit is a frequentist upper limit, which is well-known for being unreliable when there are downward fluctuations in the background (see Sec. 6.6 for further discussion), in which case the limit depend strongly on the method used to derive it.

EXO and KamLAND-Zen use ^{136}Xe as the decaying nucleus, which has its own NME, also with a large uncertainty. In order to combine and compare these results, one first needs to chose a specific underlying mechanism predicting the decay, but also take into account the statistical uncertainties and the uncertainties on both the NMEs. Due to the this difficulty, the uncertainties associated with the data used to claim an observation in ^{76}Ge, and perhaps also due to the lack of appropriate statistical tools, neither the EXO nor KamLAND-Zen collaborations attempted to really test the consistency of their data with that of the claim (not even for fixed values of the NMEs).

In paper V [5], I performed a global fit of the claim using ^{76}Ge and the recent data using ^{136}Xe within the standard model with massive Majorana neutrinos. In particular, I was interested in testing the consistency of the different data sets. Furthermore, instead of simply performing the analysis for fixed values of the NMEs, I took into account the theoretical uncertainties of the NMEs which is possible within a Bayesian analysis (see Sec. 6).

Finally, we mention that the GERDA experiment [88] is currently running. Since it uses ^{76}Ge as source, it will enable a more direct comparison with the positive claim, i.e., without any uncertainty from the NMEs. The GERDA collaboration recently released their first results on $0\nu\beta\beta$ [89]. The data show no signs of $0\nu\beta\beta$, and seem to disfavour the previous claim.

3.3.1 Other mechanisms of neutrinoless double beta decay

Although the standard mechanism for $0\nu\beta\beta$ is the exchange of light Majorana neutrinos, other mechanisms could very well appear in certain extensions of the SM, such as supersymmetric models and models with heavy neutrinos [90–93], as well as left-right symmetric models [94]. To discriminate between different mechanisms of $0\nu\beta\beta$, it will obviously not be enough to detect $0\nu\beta\beta$ in a single isotope. However, since the NMEs for different mediating mechanisms are different, one could in principle tell them apart using measurements of several different isotopes [92,95,96].

Instead of treating specific high-energy models, one can use effective field theory and look at the most general operators which could be responsible for $0\nu\beta\beta$. First, one can alter the Lorentz and/or chirality structure of the effective four-fermion interaction, while keeping the neutrino propagator [77]. There is also the possibility that no neutrino exchange is involved in the new decay mechanism. If the virtual particles responsible for the decay are heavier than the typical nuclear energies, the whole process should be describable by a single point-like interaction in effective field theory. Since in this case there are four quark fields and two electron fields involved, the effective operator has to have a mass dimension equal to nine. The most general such Lagrangian is given by [97]

$$\mathcal{L}_{0\nu\beta\beta} = \frac{G_{\mathrm{F}}^2}{2} m_{\mathrm{p}}^{-1} \left(\epsilon_1 JJj + \epsilon_2 J^{\mu\nu} J_{\mu\nu} j + \epsilon_3 J^{\mu} J_{\mu} j + \epsilon_4 J^{\mu} J_{\mu\nu} j^{\nu} + \epsilon_5 J^{\mu} J j_{\mu}\right) + \mathrm{H.c.}, \tag{3.13}$$

where J and j denote hadron and electron currents, respectively. The proportionality to the Fermi constant G_{F}^2 has been introduced, since this also appears in the standard mechanism, while the factor m_{p}^{-1} finally gives the coefficient the correct mass dimension.[2] The strengths of the different operators are parametrized by the (generally complex) dimensionless coefficients ϵ_i. Actually, there are many more

[2]The choice of the proton mass is arbitrary, but in some sense natural since the proton appears as a final state, and since the typical energy transfer inside the nucleus is substantially smaller.

operators in Eq. (3.13) then there seems to be at first sight. This is because different chirality structures are permitted for all the currents. The hadron currents in Eq. (3.13) are given by

$$J_{\mathrm{L,R}} = \overline{u}\,(1 \mp \gamma_5)\,d, \quad J^{\mu}_{\mathrm{L,R}} = \overline{u}\gamma^{\mu}\,(1 \mp \gamma_5)\,d, \quad J^{\mu\nu}_{\mathrm{L,R}} = \overline{u}\frac{i}{2}\left[\gamma^{\mu},\gamma^{\nu}\right](1 \mp \gamma_5)\,d, \tag{3.14}$$

and the electron ones by

$$j_{\mathrm{L,R}} = \overline{e}\,(1 \mp \gamma_5)\,e^{c} = 2\,\overline{e_{\mathrm{R,L}}}e^{c}_{\mathrm{R,L}}, \quad j^{\mu}_{\mathrm{L,R}} = \overline{e}\gamma^{\mu}\,(1 \mp \gamma_5)\,e^{c} = 2\,\overline{e_{\mathrm{L,R}}}\gamma^{\mu}e^{c}_{\mathrm{R,L}}. \tag{3.15}$$

Note that there are some more Lorentz invariant terms, which could have been added to Eq. (3.13), namely

$$\mathcal{L}'_{0\nu\beta\beta} = \frac{G_{\mathrm{F}}^{2}}{2}m_{\mathrm{p}}^{-1}\left(\epsilon_6 J^{\mu}J^{\nu}j_{\mu\nu} + \epsilon_7 J J^{\mu\nu}j_{\mu\nu} + \epsilon_8 J_{\mu\alpha}J^{\nu\alpha}j^{\mu}_{\nu}\right), \tag{3.16}$$

where the electron tensor currents are given by $j^{\mu\nu}_{\mathrm{L,R}} = \overline{e}\frac{i}{2}\left[\gamma^{\mu},\gamma^{\nu}\right](1 \mp \gamma_5)\,e^{c}$. However, one can show that all operators proportional to $\overline{e}\gamma^{\mu}e^{c}$, $\overline{e}\frac{i}{2}\left[\gamma^{\mu},\gamma^{\nu}\right]e^{c}$, and $\overline{e}\gamma_5\frac{i}{2}\left[\gamma^{\mu},\gamma^{\nu}\right]e^{c}$ vanish identically, since the electron fields anti-commute [98]. Thus, the terms in Eq. (3.16) all vanish and do not need to be considered.

The resulting half-life, including interferences, can be calculated from the interactions in the Lagrangian in Eq. (3.13) [97], and also interferences with the standard light neutrino exchange mechanism can be derived [3]. Note that, depending on the chiralities of the final state electrons, several of the interference terms will be suppressed, see Ref. [3] for a discussion.

Paper III of this thesis [3] deals with prospective constraints on the effective operators in Eq. (3.13) from future data on $0\nu\beta\beta$, in combination with data from single beta decay experiments and cosmological observations.

Chapter 4

Seesaw models

Seesaw models are a group of models involving new heavy degrees of freedom such that, in the low-energy theory where the heavy fields are integrated out, the effective operator in Eq. (2.45) is generated, resulting in a Majorana mass matrix for the light neutrinos. In other words, they are simple extensions of the SM such that, at low energy, the SM with the Weinberg operator is recovered and the matrix κ can be given in terms of the parameters of the high-energy theory. The Weinberg operator is usually generated at tree-level, but can also appear due to radiative corrections [99]. For the tree-level case, there are three main type of models, depending on which type of fields generate the Weinberg operator,

- *Type I* seesaw models [100–103], where a number of fermionic SM singlets, basically right-handed neutrinos, are introduced,
- *Type II* seesaw models [104–109], where scalar $\mathrm{SU(2)_L}$ triplets are introduced,
- *Type III* seesaw models [110], where fermionic $\mathrm{SU(2)_L}$ triplets are introduced.

In general, there is nothing that prevents more than one of these sets of fields to be present simultaneously, giving combinations of seesaw models.

Usually, the new fields introduced have masses far above the electroweak scale, outside the reach of any foreseeable experiments, making these versions of seesaw models essentially untestable.[1] However, there are also seesaw models where the new particles have masses above the electroweak scale, but within the reach of future experiments such as the LHC, so-called *low-scale seesaw models*. For potential collider signatures of such models, see Refs. [118, 119] and references therein.

In this chapter, the type I seesaw model as well as its variation the *inverse seesaw model* will be discussed in more detail. Both of these models can be constructed such that the new particles have masses at low energy scales, e.g., at the TeV scale,

[1] They could affect processes at very high energies. For example, they could generate the baryon asymmetry of the Universe through *leptogenesis* [111–117].

making them, in principle, testable in future experiments. The reader is referred to the references for more details on the other types of seesaw models. The type I seesaw model was studied in paper I of this thesis [1] and the inverse seesaw model in paper II [2].

4.1 The type I seesaw model

The type I seesaw model is the canonical seesaw model, and is basically a special case of the model introduced in Sec. 2.7.2, i.e., the particle content of the SM is extended with three right-handed neutrino fields $\nu_{\mathrm{R}i}$ with a Majorana mass matrix M_{R}, which has eigenvalues above the electroweak scale. Also, it is usually assumed that there are no other contributions to the masses of the light neutrinos. In this case, at energies below M_{R}, the Weinberg operator with

$$\kappa = Y_\nu M_{\mathrm{R}}^{-1} Y_\nu^T \tag{4.1}$$

is generated at tree-level.[2] This can be represented diagrammatically as in Fig. 4.1; at low energies the propagators of the right-handed neutrinos can be "shrunk" to a single point. Thus, after electroweak symmetry breaking, there is a Majorana mass term with mass matrix

$$M_{\mathrm{L}} = v^2 \kappa = v^2 \cdot Y_\nu M_{\mathrm{R}}^{-1} Y_\nu^T = F M_{\mathrm{R}} F^T, \tag{4.2}$$

with $F = v Y_\nu M_{\mathrm{R}}^{-1}$. It is thus suppressed by a factor of $Y_\nu v M_{\mathrm{R}}^{-1}$ with respect to that expected for a Dirac mass, i.e., $Y_\nu v$.

The next operator generated in the tower of effective interactions, relevant for neutrinos, is the dimension-six operator [120–123]

$$\mathcal{L}_\nu^{d=6} = \left(\overline{\ell_{\mathrm{L}}}\tilde{\phi}\right) C i\partial\!\!\!/ \left(\tilde{\phi}^\dagger \ell_{\mathrm{L}}\right), \tag{4.3}$$

where the coefficient matrix is given, at leading order, by

$$C = \left(Y_\nu M_{\mathrm{R}}^{-1}\right)\left(Y_\nu M_{\mathrm{R}}^{-1}\right)^\dagger. \tag{4.4}$$

After electroweak symmetry breaking, this dimension-six operator leads to corrections to the kinetic terms for the light neutrinos. In order to keep the neutrino kinetic energy canonically normalized, one has to rescale the neutrino fields, resulting in a non-unitary matrix relating the flavor and mass eigenstates, given by

$$N = \left(1 - \frac{v^2}{2} C\right) U = \left(1 - \frac{F F^\dagger}{2}\right) U, \tag{4.5}$$

where U diagonalizes the light neutrino mass matrix. For $|F| \gtrsim \mathcal{O}(0.1)$, non-negligible non-unitarity effects could be visible in the near detector of a future

[2] This is accurate for energies E below all the eigenvalues of M_{R}. For energies between two eigenvalues of M_{R}, only the right-handed neutrinos with masses above E should be integrated out. This will be discussed in more detail in Sec. 5.3.

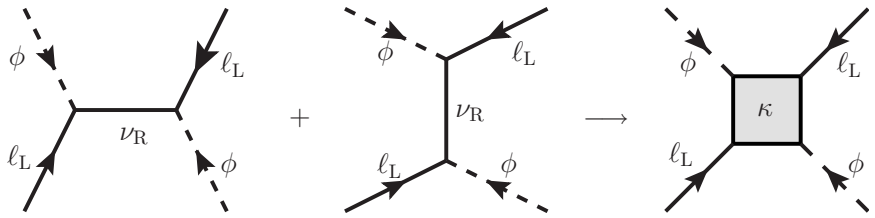

Figure 4.1. The generation of the Weinberg operator in the type I seesaw model.

neutrino factory [124–128], and there are further constraints coming from the universality tests of weak interactions, rare lepton decays, the invisible Z width, and neutrino oscillation data [129]. However, note that such a large value of F is incompatible with Eq. (4.2) and the current bounds on neutrino masses unless severe fine tuning is involved so that large cancellations occur.

The model can also be analyzed by keeping the right-handed fields in the theory, breaking electroweak symmetry spontaneously, and then approximately block diagonalizing the resulting full mass matrix, given in Eq. (2.57) with $M_{\rm L} = 0$. This can be done using

$$U \simeq \begin{pmatrix} \mathbb{1} & \rho \\ -\rho^\dagger & \mathbb{1} \end{pmatrix}, \tag{4.6}$$

with $\rho = M_{\rm D} M_{\rm R}^{-1}$, giving

$$U^T \mathcal{M}_\nu U \simeq \begin{pmatrix} -M_{\rm D} M_{\rm R}^{-1} M_{\rm D}^T & 0 \\ 0 & M_{\rm R} \end{pmatrix} = \mathcal{D}_\nu \tag{4.7}$$

to lowest order in $M_{\rm D} M_{\rm R}^{-1}$. The upper left 3×3 block is then a Majorana mass matrix for the fields $R = \nu_{\rm L} + \rho \nu_{\rm R}^c$, containing a small admixture of the gauge singlet right-handed neutrinos, proportional to $\rho \ll 1$. Also, the heavy neutrino mass eigenstate fields, which are mainly composed of the gauge singlet right-handed fields, also contain a small component of left-handed neutrino fields. Note that the matrix which enters into the lepton mixing matrix is the matrix which diagonalizes the upper left 3×3 block of the full mass matrix $\mathcal{M}_\nu$. However, this matrix will not necessarily be unitary, as it is only a part of the full unitary 6×6 matrix which diagonalizes the full mass matrix. This is how the non-unitarity enters in this way of looking at the model, which is to be compared with the effects of the dimension-six operator in Eq. (4.3).

In paper II of this thesis [2], some properties of the *low-scale* type I seesaw model were considered, in which the right-handed neutrinos have masses close to the electroweak scale. In the ordinary seesaw models with right-handed neutrino masses far above the electroweak scale, Y_ν can be sizable, e.g., at order unity. In the low-scale seesaw model, Y_ν should be relatively small in order to maintain the stability of the masses of the light neutrinos. However, there exist mechanisms that

could stabilize neutrino masses without the requirement of a tiny Y_ν. For example, additional suppression could enter through a small lepton number violating contribution, as in the inverse seesaw model (see Sec. 4.2). Also, the neutrino masses could be generated radiatively, in which case the additional suppression is guaranteed by loop integrals [99]. Finally, neutrino masses could be forbidden at $d = 5$, but appear from effective operators of higher dimension [34]. In these cases, there will still be restrictions on Y_ν from the unitarity of the lepton mixing matrix.

To fully specify the type I seesaw model one needs to specify 18 parameters, in addition to the ones in the SM [130]. This can be done in different ways.[3] For example, in the *top-down* parametrization, the model is considered at high-energy scales, where the right-handed neutrinos are propagating degrees of freedom. As mentioned before, one can always choose a $\nu_{\rm R}$ basis where the mass matrix $M_{\rm R}$ is diagonal, with positive and real eigenvalues, i.e., $M_{\rm R} = D_{\rm R}$. The remaining neutrino Yukawa matrix Y_ν is an arbitrary complex matrix, from which 3 phases can be removed by phase redefinitions of the $\ell_{{\rm L}i}$'s, giving 15 additional parameters. Another useful and popular parametrization, more natural and relevant for low-energy physics, is the *Casas–Ibarra* parametrization [131]. First, it uses the real and diagonal matrices $D_{\rm R}$, $D_\kappa = D_{\rm L}/v^2$, and the lepton mixing matrix U, containing a total of 12 parameters. The remaining 6 parameters are encoded in the matrix

$$O \equiv D_\kappa^{-1/2} U^\dagger Y_\nu M_{\rm R}^{-1/2}. \tag{4.8}$$

If the relation in Eq. (4.1) is to hold, O has to be a complex orthogonal matrix, which means that it can be written in the form $O = R_{23}(\vartheta_1) R_{13}(\vartheta_2) R_{12}(\vartheta_3)$ with $R_{ij}(\vartheta_k)$ being the elementary rotations in the 23, 13, and 12 planes, respectively. Different from the quark or lepton mixing angles, ϑ_i are in general complex.

4.2 The inverse seesaw model

In the inverse seesaw model [132], the smallness of the neutrino masses is guaranteed by a small amount of lepton number breaking instead of suppression by a very large mass scale. It contains three extra fermionic SM gauge singlets S_i, coupled to the right-handed neutrinos in a lepton-number conserving way, while the ordinary right-handed neutrino Majorana mass term is forbidden by some additional symmetry. It is only through a symmetric mass matrix $M_{\rm S}$ in the Majorana mass term $\overline{S^c} M_{\rm S} S$ that the lepton number is broken, and $M_{\rm S}$ can thus be naturally small. The relevant part of the Lagrangian is then, in the flavor basis,

$$-\mathcal{L}_{\rm IS} = \overline{\ell_{\rm L}} \tilde{\phi} Y_\nu \nu_{\rm R} + \overline{S^c} M_{\rm R} \nu_{\rm R} + \frac{1}{2} \overline{S^c} M_{\rm S} S + \text{H.c.} \tag{4.9}$$

[3] One can always choose the $\ell_{\rm L}$ basis such that $Y_e^\dagger Y_e = D_e$, containing three parameters.

Here, the fields $\nu_{\mathrm{R}i}$ and S_i are not mass eigenstates, but instead the Majorana mass matrix in the basis $\{\nu_{\mathrm{R}}, S\}$ is

$$\mathcal{M}_{\mathrm{IS}} = \begin{pmatrix} 0 & M_{\mathrm{R}} \\ M_{\mathrm{R}}^T & M_{\mathrm{S}} \end{pmatrix}. \tag{4.10}$$

For $M_{\mathrm{S}} \ll M_{\mathrm{R}}$, the right-handed neutrinos and the extra singlets S_i are, to lowest order, maximally mixed into three pairs of heavy Majorana neutrinos with opposite CP parities and essentially identical masses, with a splitting of the order of M_{S}, and can as such be regarded as components of three heavy pseudo-Dirac neutrinos.

Integrating out these heavy fields yields the Weinberg operator with

$$\kappa = \left(Y_\nu M_{\mathrm{R}}^{-1}\right) M_{\mathrm{S}} \left(Y_\nu M_{\mathrm{R}}^{-1}\right)^T \tag{4.11}$$

at tree-level, which, after electroweak symmetry breaking as usual, yields a Majorana mass matrix for the light neutrinos as

$$M_{\mathrm{L}} = F M_{\mathrm{S}} F^T, \tag{4.12}$$

where $F = v Y_\nu M_{\mathrm{R}}^{-1}$. This is to be compared with Eq. (4.2) for the type I seesaw model. The diagrammatical representation is still given by the diagrams in Fig. 4.1, but with all the 6 heavy mass eigenstate fields appearing as intermediate states.

In spite of the underlying physics responsible, the particle content of the inverse seesaw model is essentially the same as that of the type I seesaw model, but with six right-handed neutrinos. Thus, one can in principle treat the heavy singlets S_i as three additional right-handed neutrinos, possessing vanishing Yukawa couplings with the lepton doublets. It is also worth comparing the canonical type I and inverse seesaw models with the discussion of the Weinberg operator in Sec. 2.7.1. In the seesaw models, the cutoff scale Λ_ν in Eq. (2.51) can essentially be identified with M_{R}, which is generally above the electroweak scale. However, the dimensionless $\tilde{\kappa}$ in Eqs. (2.51) and (2.52) then have the order of magnitudes

$$\tilde{\kappa} = \begin{cases} \mathcal{O}(Y_\nu^2) & \text{canonical type I seesaw,} \\ \mathcal{O}(Y_\nu^2 M_{\mathrm{S}} M_R^{-1}) & \text{inverse seesaw.} \end{cases} \tag{4.13}$$

Thus, $\tilde{\kappa}$ can be strongly suppressed by the potentially very small ratio $M_{\mathrm{S}} M_R^{-1}$ in the inverse seesaw model.

Finally, note that, in the inverse seesaw model, the correct light neutrino masses can be obtained even for $F = \mathcal{O}(1)$, i.e., for the new heavy fields around the electroweak scale and with large Yukawa couplings Y_ν, and that the non-unitarity effects are, as in the type I seesaw model, given by Eq. (4.5). Thus, as opposed to the ordinary type I seesaw model, large non-unitarity effects are possible in the inverse seesaw model.

Chapter 5

Renormalization group running

This chapter is a description of the concept of renormalization group (RG) running. The need for regularization and renormalization is described using a simple example and the motivations for studying the RG running in the SM and seesaw models as well as the methods for solving the resulting RG equations are reviewed. The decoupling of the right-handed neutrinos and the use of effective theory is explained. Finally, the proper description of the running between the masses of the heavy particles and how this can lead to the so-called threshold effects in the running of the neutrino parameters is described.

5.1 The main idea

Calculations of quantum corrections, represented by loops in Feynman diagrams, to physical quantities (such as cross sections, decay rates, and particle masses), as well as unphysical ones (such as correlation functions), often yield divergent results. This implies that the calculated corrections are not uniquely defined, and as a result, neither are the predictions of the theory.

The standard way to deal with this issue is to implement a two-step procedure. First, one has to *regularize* the divergence by modifying the theory in some way. This is performed by introducing some parameter ϵ, such that the modified prediction is a well-defined function of ϵ and the original, divergent result is reobtained in the limit $\epsilon \to 0$. Then, one has to *renormalize* the theory by redefining its parameters, such that the prediction becomes finite in the $\epsilon \to 0$ limit. For this to be the case, the original parameters and fields appearing in the Lagrangian, the so-called *bare* parameters and fields, must formally diverge as $\epsilon \to 0$. In order to make these concepts easier to grasp, a simple example will be used for illustration.

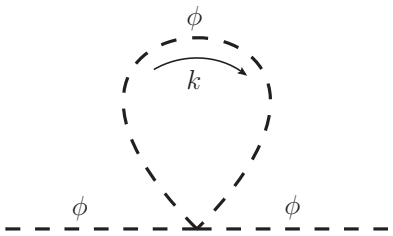

Figure 5.1. Self-energy diagram of a scalar field ϕ. Here, k is the loop momentum.

Consider the QFT with only a single real scalar field ϕ with mass m and quartic self-coupling λ. The one-loop self-energy diagram is given in Fig. 5.1, the value of which is

$$\mathrm{i}\Sigma_\phi = \frac{\lambda}{2}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{k^2 - m^2 + \mathrm{i}\epsilon}, \tag{5.1}$$

where k is the loop momentum. One way to evaluate these kinds of integrals is to perform a *Wick rotation* by changing variables to

$$k^0 \equiv \mathrm{i}k^4_\mathrm{E}, \quad \boldsymbol{k} \equiv \boldsymbol{k}_\mathrm{E}, \tag{5.2}$$

which implies that the Lorentz inner product is given by

$$k^2 = (k^0)^2 - \boldsymbol{k}^2 = -(k^4_\mathrm{E})^2 - \boldsymbol{k}^2_\mathrm{E} = -k^2_\mathrm{E}, \tag{5.3}$$

where $k^2_\mathrm{E} = (k^4_\mathrm{E})^2 + \boldsymbol{k}^2_\mathrm{E}$ is just the ordinary inner product in four-dimensional Euclidean space. Re-routing the integral in the complex plane and then going to spherical coordinates, $\mathrm{i}\Sigma_\phi$ can be calculated as

$$\mathrm{i}\Sigma_\phi = -\mathrm{i}\frac{\lambda}{2(2\pi)^4}\int \mathrm{d}\Omega \int_0^\infty \mathrm{d}k_\mathrm{E}\frac{k^3_\mathrm{E}}{k^2_\mathrm{E} + m^2}, \tag{5.4}$$

which is divergent in the region of large k_E.

It is now time to regularize this integral, which in general can be done in a number of different ways. The simplest way, and arguably the physically most intuitive, is to use an ultraviolet cutoff. Simply cut off the integral at some large energy scale $k_\mathrm{E} = \Lambda$, i.e., only integrate up to Λ instead of ∞, giving

$$\Sigma_\phi = -\frac{\lambda}{32\pi^2}\left[\Lambda^2 - m^2\log\left(1 + \frac{\Lambda^2}{m^2}\right)\right]. \tag{5.5}$$

The parameter ϵ can then, for example, be chosen as $\epsilon = \Lambda^{-1}$. Note that Λ is not the (fixed) energy scale up to which your theory is valid, but an arbitrary

regularization scale. This regularization method has the disadvantage that is breaks gauge invariance, and is thus not suitable for gauge theories.

Another regularization method is the *Pauli-Villars* regularization, in which a smooth cutoff is introduced in the propagator by making the replacement

$$\frac{\mathrm{i}}{k^2-m^2} \to \frac{\mathrm{i}}{k^2-m^2} - \frac{\mathrm{i}}{k^2-M^2} = \frac{\mathrm{i}(m^2-M^2)}{(k^2-m^2)(k^2-M^2)}, \tag{5.6}$$

which can also be viewed as the introduction of a new fictitious particle with a large mass M and wrong overall sign of the propagator. The parameter ϵ can be taken to be M^{-1}. However, this method becomes too complicated for less simple theories, such as the SM.

Lattice regularization implies replacing the space-time continuum by a lattice with finite spacing $\epsilon = l$, removing modes of the field with momenta larger than l^{-1}. This regularization is automatically present in numerical non-perturbative calculations, but less suitable for analytical calculations using perturbation theory, since this regulator breaks Lorentz invariance.

Finally, the most widely used regularization method, which preserves Lorentz and gauge invariance, but is perhaps the most non-intuitive one, is *dimensional regularization*. Here, the number of space-time dimensions is altered to $d = 4 - \epsilon$. Of course, one has to make sure that the mathematical framework one is using is properly generalized to arbitrary d. For example, the mass dimensions of the fields have now changed, so that Eqs. (2.29) and (2.30) are generalized to

$$[\phi] = [A^\mu] = \frac{d-2}{2}, \tag{5.7}$$

$$[\psi] = \frac{d-1}{2}. \tag{5.8}$$

In general, all the mass parameters still have the dimension of a mass, but all the other coupling constants need to be redefined in order to keep their mass dimensions (they are typically dimensionless). For example, in the SM and related theories, one has to make the replacements

$$\lambda \to \lambda_0 = \mu^\epsilon \lambda, \tag{5.9}$$

$$g_i \to g_{i0} = \mu^{\frac{\epsilon}{2}} g_i \quad \text{for } i \in \{1,2,3\}, \tag{5.10}$$

$$Y_f \to Y_{f0} = \mu^{\frac{\epsilon}{2}} Y_f \quad \text{for } f \in \{u,d,\nu,e\}. \tag{5.11}$$

Here, μ is an arbitrary energy scale, called the *renormalization scale*, and the subscript "0" denotes the bare quantities which have mass dimensions, while the couplings without this subscript denote the renormalized couplings. These relations between the bare and renormalized couplings are only the lowest order results, while loop corrections will modify the relations to include the renormalization constants.

Continuing the example, by Wick rotating and using standard formulae [8], the scalar self-energy can then be calculated as

$$\begin{aligned} \mathrm{i}\Sigma_\phi &= \mu^\epsilon \frac{\lambda}{2} \int \frac{\mathrm{d}^d k}{(2\pi)^4} \frac{1}{k^2 - m^2 + \mathrm{i}\epsilon} \\ &= -\mathrm{i}\frac{\lambda}{32\pi^2} (4\pi)^{\frac{\epsilon}{2}} \Gamma\left(-1 + \frac{\epsilon}{2}\right) m^2 \left(\frac{\mu^2}{m^2}\right)^{\frac{\epsilon}{2}} \\ &= \mathrm{i}\frac{\lambda}{32\pi^2} m^2 \left[\frac{2}{\epsilon} + \log\frac{\mu^2}{m^2} + \log 4\pi + 1 - \gamma_\mathrm{E} + \mathcal{O}(\epsilon)\right], \end{aligned} \tag{5.12}$$

where the expression in the middle line has been expanded in ϵ to yield the last line, and $\gamma_\mathrm{E} \simeq -0.5572$ is Euler's constant. It is now clear in which way the integral diverges in the limit $\epsilon \to 0$, and how the introduction of the scale μ ensures that the argument of the logarithm is dimensionless.

After having regularized a loop diagram, it is time for the step of renormalization. In general, loop corrections to two-point functions, as the one in the example, yield corrections to the corresponding field's mass and wave function normalization. If the corresponding correction is divergent in the limit $\epsilon \to 0$, the mass and wave function renormalizations also have to be divergent in this limit, leaving the renormalized masses and fields finite. Loop corrections to higher-order correlation functions require renormalization of the corresponding coupling constants. For example, a loop diagram with four external scalars requires renormalization of the coupling constant λ_0, yielding the renormalized coupling. There are in general a number of ways of renormalizing a QFT, the main groups of *renormalization schemes* being the *mass-dependent* and *mass-independent* schemes, meaning that the *counterterms* introduced to cancel the divergences are dependent and independent of the mass parameters of the theory, respectively.

The renormalized parameters are the ones one should relate to experiments, although they are not observables in the strict sense. For example, in perturbation theory, predictions for observables are expansions in the renormalized couplings, which are functions of the renormalization scale. Fixing a renormalization scale (usually of the order of the relevant energy in the process), the values of the coupling constants at that renormalization scale can be inferred from the experimental data. Note that the exact result for a physical observable should be independent of the renormalization scale (and more generally, the renormalization scheme), while individual terms in the perturbation expansion are not necessarily so. Thus, by choosing the renormalization scheme and scale wisely, i.e., in a way that the effective expansion parameter becomes small, one can optimize the expansion.

Writing m_0 for the bare mass of our example scalar field, the corrected mass squared will be $m_0^2 - \Sigma_\phi$.[1] Defining the counterterm δm^2 by $m_0^2 \equiv m^2 + \delta m^2$, the corrected propagator and the renormalized mass m will be finite if δm^2 is made

[1] In general, there will be a wave function renormalization as well, but in this example there is no need for this.

to diverge in such a way so that it cancels the divergence of Σ_ϕ exactly. There are different ways of accomplishing this, corresponding to different renormalization schemes, but the most widely used is the Minimal Subtraction (MS) scheme, or its modified version $\overline{\text{MS}}$. In MS, only the poles in ϵ are subtracted.[2] In our example, one takes

$$\delta m^2 = \frac{\lambda}{16\pi^2} m^2 \frac{1}{\epsilon}. \tag{5.13}$$

Now, by noting that all the bare parameters are independent of the renormalization scale μ, one can from Eq. (5.9) calculate that

$$\mu \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = -\epsilon\lambda, \tag{5.14}$$

and hence

$$\mu \frac{\mathrm{d}m^2}{\mathrm{d}\mu} = \frac{\lambda}{16\pi^2} m^2, \tag{5.15}$$

and so the renormalized mass depends on the renormalization scale. Note that m^2 is renormalized multiplicatively, implying that if $m = 0$ at tree level, it will remain so after quantum effects are considered.

The above statement is generally true for fermion fields, but *not* for scalar fields. This is because the fermion mass terms generally break chiral symmetry, implying that it is natural to have them small, since the symmetry of the theory is then increased by setting the masses to zero. This is usually not the case for scalar masses, and so, using dimensional regularization, the quantum corrections to the SM Higgs mass are in general proportional to the mass of any particle running in the loop (but not to a high-energy cutoff scale). This phenomenon is often called the *hierarchy problem*, since it is argued that in order to obtain the small value of the Higgs mass observed at low energies, a large degree of "fine-tuning" is necessary. As an example, in the type I seesaw model there are corrections to the Higgs mass proportional to the Majorana masses of the right-handed neutrinos, which are generally much larger than the electroweak scale. Thus, it is not natural to have the SM Higgs mass at the electroweak scale. However, in the (unbroken) SM, where there are no right-handed neutrinos, all fermions are massless, and thus all quantum corrections to the Higgs mass are proportional to the mass itself, and thus smaller than the tree-level mass.[3]

The hierarchy problem is thus not really a problem in the SM, but rather a potential problem of high-energy extensions of the SM, in which particles with masses far above the electroweak scale appear. In these theories, obtaining a Higgs mass at the electroweak scale can involve fine-tuning between different parameters. See Ref. [133] for a recent discussion. This fine-tuning problem can be avoided if,

[2] In $\overline{\text{MS}}$, also the constant term $\ln 4\pi - \gamma_\text{E}$ is subtracted, generally leading to a better convergence of the perturbation series.

[3] This is true even if one considers the broken SM with massive fermions, since all fermions have masses at the electroweak scale or below.

for example, there are additional symmetries in the high-energy theory protecting the Higgs mass, or if the Higgs is a composite particle.

5.2 Renormalization group running of neutrino parameters in seesaw models

The example in the previous chapter can be generalized to more complicated theories, such as the SM and the seesaw models. The resulting *renormalization group equations* (RGEs) describe the dependence of the renormalized parameters on the renormalization scale μ, and constitute a system of coupled ordinary differential equations, one for each parameter. In mass-independent renormalization schemes, the RGEs have the general form

$$\mu \frac{\mathrm{d}P_i}{\mathrm{d}\mu} \equiv \beta_i \left(P_1, \ldots, P_n\right), \tag{5.16}$$

where P_i $(i = 1, \ldots, n)$ are the n parameters of the theory and the *beta functions* β_i do not depend explicitly on μ. The RGEs for the type I, type II, and type III seesaw models with both the SM and the *minimal supersymmetric standard model* (MSSM)[4] as underlying theories have been derived in the literature [137–143]. In this chapter, mainly the RGEs of the type I seesaw model will be discussed. They can be found in Appendix A, both for the extended SM and MSSM. The RGEs of both the SM and the inverse seesaw model can be obtained as special cases of the ones for the type I seesaw model.

Although it can be interesting to study the running of the gauge couplings and quark masses (or, rather, the quark Yukawa couplings), we are here mainly interested in is the RG evolution of the parameters in the lepton sector of the SM, and more specifically, the light neutrino masses and the lepton mixing parameters. In the case of the parameters outside the lepton sector of the SM, the RG running is usually studied because of the need to reconcile experimental measurements at different energies, which without any running would not be compatible. For example, this is the case for the electromagnetic coupling constant. However, this is not the case in the lepton sector, since the current experimental uncertainties are generally much larger than the running effects, and because the lepton parameters have only been measured at relatively low energies so far. Instead, the reason to study them is that theoretical predictions of models beyond the SM, such as *grand unified theories* (GUTs), are valid at some high-energy scale, while experimental data are taken at low energies. Therefore, one has to take into account the running of the parameters between the high-energy (GUT) and low-energy (experimental) scales in order to compare the experimental results with the theoretical predictions. Extensions of the SM sometimes predict specific mixing patterns in the lepton sector, i.e., specific

[4]The MSSM is a extensively studied extension of the SM, where every SM particle has an additional partner having spin differing by one half. See, for example, Refs. [134–136] for reviews.

values for the lepton mixing matrix. Two such common symmetric mixing patters are the *bimaximal* mixing pattern [144–147] with $s_{12} = s_{23} = 1/\sqrt{2}$ and $s_{13} = 0$ [see Eq. (2.50)] and the *tri-bimaximal* mixing pattern [148–150] with $s_{12} = 1/\sqrt{3}$, $s_{23} = 1/\sqrt{2}$, and $s_{13} = 0$.

The RGE evolution of the neutrino masses and lepton mixing parameters can be determined through the evolution of the effective light neutrino mass matrix (in the effective or full theories) and the charged lepton Yukawa matrix, see Sec. 2.7.1. Often, the RG running is calculated numerically, after which the neutrino masses and lepton mixing parameters are determined by diagonalizing the mass matrices. However, one can also translate the full RGEs for the neutrino mass matrix into a system of differential equations for mixing angles, CP-violating phases, and light neutrino masses directly. The corresponding formulas have been discussed below the seesaw scale [151–153], as well as above the seesaw thresholds in the type I [154, 155], type II [141, 142], and type III [143] seesaw frameworks. Note that the usual diagonality assumption made on Y_e is not in general invariant under the RG running, and neither is the parametrization of the lepton mixing matrix in Eq. (2.50), which generally includes three additional unphysical phases which have to be rotated away in order to determine the physical mixing parameters.

There are two different strategies for solving the RGEs. In the *top-down* approach, the initial conditions on the parameters are specified at a certain high-energy scale, often motivated by the flavor structure of a specific high-energy model. Once this is done, the running down to low energies and crossing the seesaw thresholds is relatively straightforward. In this approach, the main issue is the fact that only small regions of the parameter space of the full theory will lead to values of the low-energy parameters that are consistent with experiments, and this makes this approach difficult to implement in practice. In the *bottom-up* approach, on the other hand, the initial conditions on the parameters are specified at a low-energy scale, usually the electroweak scale. Hence, all the available experimental information is taken into account from the start. However, after running to higher energy scales, one reaches the seesaw threshold, where one has to match the effective and full theories. Then, since the number of parameters in the full theory is larger than in the effective one, one has to make additional assumptions on the parameters and flavor structure of the full theory.

The general features of the running of the lepton parameters have been studied in the literature, and it has been shown that there could be large radiative corrections to the lepton mixing parameters at very high energy scales (see, e.g., Ref. [156] and references therein). In particular, certain flavor symmetric mixing patterns can be achieved at the GUT scale indicating that there might exist some flavor symmetries similar to the gauge symmetry (see, e.g., Ref. [157] and references therein).

5.3 Decoupling of right-handed neutrinos and threshold effects

The description using effective field theory plays an essential role in the study of the RG running of QFT parameters. The *Appelquist–Carazzone* theorem [158] states that the effect of heavy particles decouples at energies much smaller than their masses, and that they do not contribute to the beta functions at low energies. This can be seen explicitly if one uses a mass-dependent renormalization scheme, such as momentum space subtraction. However, this does not happen if one uses a mass-independent renormalization scheme, such as MS, since the beta functions are independent of masses. Generally, in perturbative calculations, one will obtain finite contributions to observables on the form

$$\log \frac{E^2}{\mu^2},$$

which is the reason why one generally should take $\mu \simeq E$ to minimize the effects of higher-order terms. However, one can also have potentially large logarithms,

$$\log \frac{M^2}{\mu^2},$$

where M is the mass of the heavy particle [26]. If $M \gg E$, these terms might destroy the perturbation expansion. In order to implement the decoupling in mass-independent schemes, one decouples the heavy particles "by hand" by integrating them out at the matching scale $\mu \simeq M$, and describing the RG running for $\mu < M$ using the effective theory.

This holds in general for all particles, and in particular for the ones in the SM, but here we will concentrate on the heavy neutrinos, which are assumed to have masses above the electroweak scale. In the previous discussion of the seesaw models, only the different regions of energy $E < M_{\rm R}$ and $E > M_{\rm R}$ were considered. However, the three right-handed neutrinos do in general not have the same masses, i.e., the masses can be non-degenerate. In that case, the heavy neutrinos have to be sequentially decoupled from the theory [159], leading to a series of effective field theories. Once again, it is worth pointing out that perturbative renormalization of effective operators can be performed in the usual way, as long as one is satisfied with a finite accuracy and works to a given order in E/Λ.

When crossing the seesaw thresholds, one should make sure that the full and effective theories give identical predictions for physical quantities at low-energy scales, and therefore, the physical parameters of both theories have to be related to each other. In the case of the neutrino mass matrix, this means relations between the effective coupling matrix κ and the parameters Y_ν and $M_{\rm R}$ of the full theory. This is called *matching* the full and effective theories. For the simplest case, when the mass spectrum of the heavy singlets is degenerate, namely $M_1 = M_2 = M_3 =$

M_0, one can simply make use of the tree-level matching condition at the scale $\mu = M_0$,

$$\kappa\big|_{M_0} = Y_\nu M_{\rm R}^{-1} Y_\nu^T\big|_{M_0}. \tag{5.17}$$

In the most general case with non-degenerate heavy singlets, i.e., $M_1 < M_2 < M_3$, the situation becomes more complicated. For μ between M_n and M_{n-1}, the heavy mass eigenstates $\{\nu_{\rm R}^n \dots \nu_{\rm R}^3\}$ are integrated out. In this effective theory, only a $3\times(n-1)$ sub-matrix of Y_ν remains, denoted by $Y_\nu^{(n)}$, as well as an $(n-1)\times(n-1)$ submatrix of $M_{\rm R}$, denoted by $M_{\rm R}^{(n)}$. The decoupling of the n-th heavy singlet leads to the appearance of an effective dimension-five operator through the tree-level matching condition at $\mu = M_n$,

$$\kappa^{(n)}\big|_{M_n} = \kappa^{(n+1)}\big|_{M_n} + \frac{\mathcal{Y}_\nu^{(n)}\mathcal{Y}_\nu^{(n)T}}{M_n}\Big|_{M_n}, \quad \text{for } n = 1,2,3, \tag{5.18}$$

where $\mathcal{Y}_\nu^{(n)}$ is the n-th column of Y_ν (i.e., the part which has been removed from Y_ν), and it is understood that $\kappa^{(4)} = 0$ is the effective operator in the full theory and $\kappa^{(1)} = \kappa$ is the effective operator with all the heavy fields decoupled. In between these scales, all the parameters are to be run using their respective RGEs. Note that the matching has to be done in a basis where the right-handed mass matrix is diagonal, since it is the eigenstate with a specific mass which is to be decoupled.

The renormalized effective neutrino Majorana mass matrix for μ below M_n is described by two parts as

$$m_\nu^{(n)} = v^2 \left[\kappa^{(n)} + Y_\nu^{(n)} \left(M_{\rm R}^{(n)}\right)^{-1} Y_\nu^{(n)T}\right], \tag{5.19}$$

where (n) labels the quantities relevant for the effective theory between the n-th and $(n-1)$-th thresholds. Both of these contributions run with the renormalization scale μ, and the running can be determined from Eqs. (A.1g), (A.1h), and (A.1j). As it turns out, the flavor non-diagonal parts of the running are the same in both the SM and the MSSM. However, the RGEs for the two terms have different flavor diagonal contributions, but only in the SM and not in the MSSM. In particular, the coefficients for the gauge coupling and Higgs self-coupling contributions are different. The flavor diagonal parts are α_κ and $2\alpha_\nu$, respectively, and they differ as

$$\alpha_\kappa - 2\alpha_\nu = \lambda + \frac{9}{10} g_1^2 + \frac{3}{2} g_2^2 \qquad \text{in the SM,} \tag{5.20}$$

$$\alpha_\kappa - 2\alpha_\mu = 0 \qquad \text{in the MSSM.} \tag{5.21}$$

Thus, the running of the two different parts contributing to the effective neutrino mass matrix in Eq. (5.19) has different gauge and Higgs self-coupling contributions. Since these couplings are in general rather large, there can be potentially large running of the lepton mixing angles due to this "mismatch" between the two contributions. These effects are usually referred to as *threshold effects*.

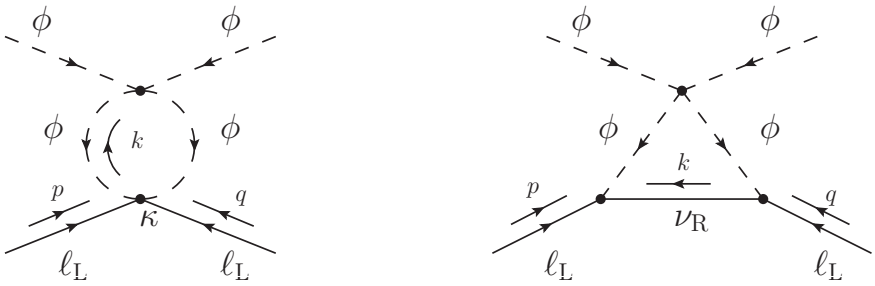

Figure 5.2. Feynman diagrams of the one-loop corrections to the neutrino mass matrix due to the Higgs self coupling in the effective theory (left) and in the full theory (right). Here, p and q are the external momenta and k is the loop momentum.

As an example to visualize where this difference comes from, consider the corrections to the four-point functions relevant to the neutrino mass matrix from the Higgs self-coupling in Fig. 5.2. In the effective theory (left diagram), there is a contribution involving the loop integral

$$\int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{k^2 - m^2} \frac{1}{(k-p-q)^2 - m^2},$$

which is divergent, while the loop integral appearing in the full theory (right diagram) is given by

$$\int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{\not{k} + M_n}{k^2 - M_n^2} \frac{1}{(k+p)^2 - m^2} \frac{1}{(q-k)^2 - m^2}$$

and is not divergent. Therefore, there are no corrections proportional to λ to the neutrino mass matrix in the full theory, while λ does enter the beta function of κ in the effective theory. See also Ref. [154] for a detailed discussion. If the relevant seesaw threshold is above the SUSY-breaking scale, such a mismatch is absent in the MSSM due to the supersymmetric structure of the MSSM Higgs and gauge sectors. Therefore, this may result in significant RG running effects only in the SM, but not in the MSSM.

In paper II of this thesis [2], the threshold effects on the RG running of the neutrino parameters in the type I seesaw model have been studied, while in paper I [1], the general RG running of the neutrino parameters in the inverse seesaw model has been investigated, including threshold effects.

Chapter 6

Statistical methods

As discussed in the introduction, science progresses through experimentation, observation, and measuring properties of the Universe around us, and then comparing the results with those predicted by theoretical models. However, no experiment is perfect. Instead, there are always uncertainties and noise which can make it difficult to compare the obtained experimental results with model predictions. This is essentially the standard case at the frontiers of physics, where one is looking for small signals or effects not previously observed. In this case, one needs to analyze the data while trying to model the present uncertainties.

There are different opinions on the usefulness of more sophisticated modelling and statistical analyses. One common viewpoint is that there is no point worrying about fine details since there will anyway in the close future be more sensitive data that will definitely find an effect if it is there, and in any case render the current experiment obsolete. Particle physics, in particular, has arguably for a long time been "spoiled" with having a constant and rapid improvement of experimental technology. However, particle physics experiments are constantly becoming more and more complex and expensive, and also the time between each generation of experiment tends to increase. The canonical example is of course the LHC, but also proposed future colliders as well as neutrino-related experiments are becoming more expensive, while offering less radical improvements over previous generations of experiments. Hence, there are many situations where there might be a long time until there will be significantly better data, and in the mean time one should try to do the best with the existing data. Additionally, as a theorist, it is desirable to have a good understanding of statistics, since a very common error among theorists is to put too much emphasis on small deviations in data without any real statistical significance.[1]

[1]However, for theorists, the potential benefit of being first to describe some feature in the data in terms of a new particle or other physical processes might be very much larger than the potential loss (as wasted time) associated with investigating a model. Hence, it can be logical to look for a

A physical model with free parameters generally leads to predictions of experimental outcomes as a function of those parameters. Here, we will only be concerned with *parametric inference*, in which one models the data using some parametrized probability distribution. Non-parametric inference is also possible, where one aims at making inference without assuming a particular model for the data. An example of this is estimating the probability distribution of some quantity by only using the observed values – a simple histogram is an example.

6.1 Probability

The field of statistics and data analysis rely heavily on the use of *probabilities* of different kinds of events and propositions. However, what is meant by "probability" and how it should be defined is not uniquely settled, and this can often cause severe confusion when attempting to perform even simple tasks of data analysis. On the contrary, there are two main interpretations of the concept of probability.

In the *frequentist* interpretation, probability describes "randomness" and is defined as the limit of the relative frequency of an event in a large number of repeated trials. For example, if a die is tossed many times and the fraction of a certain outcome (say "4") converges to a certain number p as it is tossed more and more times, then the probability of that outcome is p. In the *Bayesian* interpretation, probability describes uncertainty and is more generally defined as a quantification of the *plausibility* (or *credibility*, or *degree-of-belief*) of a proposition. This is more general since probability can be defined for essentially any proposition, not only those related to repeatable phenomena. There is not much point in arguing which definition is "correct" (as for all definitions), and they could both be internally consistent. Instead, the key issue is which definition is more useful or appropriate, and allows one to investigate the actual questions of interest.

Adopting a notation suited for the Bayesian interpretation[2], we will write $\Pr(A|I)$ for "the probability of A, given that I is true". $\overline{A}$ means "not A", while A, B means "A and B".

Although the preferred interpretation of probability has long been debated, and still is, what is meant by probability in a mathematical sense is rather well-established. Basically, probabilities are real numbers between 0 and 1 which conform to the rules

$$\Pr(A, B|I) = \Pr(A|B, I)\Pr(B|I) \tag{6.1}$$

$$\Pr(A|I) + \Pr(\overline{A}|I) = 1, \tag{6.2}$$

often called the *product rule* and *sum rule*, respectively. The first of these rules makes it possible to derive how the probabilities change when changing the ordering

possible cause of some potential anomaly in the data even when the plausibility that a new effect actually has been found is very small.

[2]Frequentist probabilities can be obtained as special cases.

of the conditioning: this is *Bayes' theorem*

$$\Pr(A|B,I) = \frac{\Pr(B|A,I)\Pr(A|I)}{\Pr(B|I)}. \tag{6.3}$$

This relation is very useful when it is applied to the case where B represents a set of collected data and A a statistical model describing how the data is generated. This will be discussed in detail in the rest of this chapter.

If a number of propositions $\{B_1, \ldots, B_n\}$ are mutually exclusive (i.e., only one of them can be true) and exhaustive (i.e., one of them must be true), then

$$\sum_{i=1}^{n} \Pr(B_i|I) = 1, \tag{6.4}$$

i.e., the total probability must sum to unity. Furthermore, one can show that

$$\Pr(A|I) = \sum_{i=1}^{n} \Pr(A, B_i|I) = \sum_{i=1}^{n} \Pr(A|B_i, I)\Pr(B_i|I). \tag{6.5}$$

This relation is very useful when one wants to calculate the probability of A but does not care about the probabilities of the B_i's. When we later study Bayesian inference, this will be useful as this allows one to effectively eliminate *nuisance parameters*, i.e., parameters of no immediate interest, when comparing different statistical hypotheses.

That frequentist probabilities, i.e., long-term relative frequencies, should follow the rules of Eqs. (6.1) and (6.2) should be relatively clear. But why should the more general concept of the plausibility of a proposition be given by a probability, i.e., why should plausibilities follow the mathematical laws of probability theory?

One of the most convincing arguments was originally pointed out by Cox [160] and beautifully described by Jaynes [161]. If one assumes a small set of reasonable *desiderata* which any measure of plausibility should have, these are enough to completely specify a theory for inference that is equivalent to probability theory. These desiderata can be summarized as follows. First, one should be able to compare all plausibilities, which means they can be represented by real numbers. Second, they should qualitatively agree with common sense. As an example, this can mean that if, after having learned some new information, the plausibility of A increases, but that of B is unaffected, then the plausibility of (A and B) must increase. Finally, one demands consistency, so that, among other things, different ways of calculating plausibilities must give the same answer. From these desiderata it is essentially possible to show that the plausibilities must satisfy the "axioms" of probability theory.[3]

[3]Technically, there are an infinite number of real consistent representations of plausibility, but any such can always be monotonously transformed to probabilities. As put by Loredo [162]: "Every allowed plausibility theory is isomorphic to probability theory. The various allowed plausibility theories may differ in form from probability theory, but not in content."

The above desiderata only refers to plausibilities of propositions, but not to frequencies, ensembles, repeated experiments, or random variables. Also, deductive logic (i.e., deducing true statements from other true statements) is a limiting case in which "0" represents falsehood and "1" represents truth. Hence, Bayesian inference can be viewed as the extension of logic to situations in which there is uncertainty, and then also the *unique* such extension satisfying the basic consistency conditions outlined above.

There also exists a motivation for Bayesian probability using the notion of the *coherent bet* of de Finetti. Roughly speaking, this means that one should assign and manipulate probabilities so that one cannot be made a sure loser in a bet based on them. However, to base inference in terms of optimizing financial gain seems rather non-scientific. For comparison with other systems of probability, such as that of de Finetti and the set-theoretic one of Kolmogorov, see Ref. [161].

In the next sections, the application of Bayesian probability to the analysis of data and inference over physical models and their parameters will be discussed. Then, we will often have "probability distributions of parameters" (and hypotheses). As an example, say we know that the value of the physical parameter $\boldsymbol{\Theta}$ must either equal $\boldsymbol{\Theta}_0$ or $\boldsymbol{\Theta}_1$. Then, we have Bayesian probabilities $\Pr(\text{“}\boldsymbol{\Theta} = \boldsymbol{\Theta}_0\text{”})$ and $\Pr(\text{“}\boldsymbol{\Theta} = \boldsymbol{\Theta}_1\text{”})$ of the respective propositions. In order to make the notation more concise, we will denote these by $\Pr(\boldsymbol{\Theta}_0)$ and $\Pr(\boldsymbol{\Theta}_1)$, but keep in mind that these are not probabilities that a parameter will take on a certain value per se (as it would be if one would naively apply the frequentist notion of probability). It is the probability that is distributed over the space of different hypotheses (and parameter values), but the hypothesis (or the parameter) is not distributed and is not "random". Each parameter has a single "true" value (or at least we can assume it has), and we want to infer that value. However, since we do not know this value exactly, we must describe this uncertainty by spreading the probability over the set of all possible values.

In this chapter we will concentrate heavily on different applications using Bayesian probabilities, and only have a shorter discussion on the most common non-Bayesian methods in Sec. 6.6.

6.2 Bayesian inference

Perhaps the main goal of science is to infer which model or hypothesis best describes a certain set of collected data. Also, as mentioned in the introduction, these models should preferably be "simple" or "economical" in some sense. If one accepts the Bayesian interpretation of probability, a very powerful arsenal of inference tools become available. In a nutshell, the idea is to use the laws of probability to calculate the probabilities (i.e., plausibilities) of different hypotheses or models, when conditioned on some known (or presumed) information.

If the collected data is denoted by $\mathbf{D}$ and the set of possible hypotheses or models is $H_1, H_2, \ldots, H_r$, the Bayesian solution is to use Bayes' theorem to calculate the

plausibilities of each of the hypotheses after considering the data, the *posterior probabilities*,

$$\Pr(H_i|\mathbf{D}) = \frac{\Pr(\mathbf{D}|H_i)\Pr(H_i)}{\Pr(\mathbf{D})}. \tag{6.6}$$

Here $\Pr(\mathbf{D}|H_i)$ is the probability of the data, assuming the model H_i to be true, while $\Pr(H_i)$ is the *prior probability* of H_i, which is how plausible H_i is before considering the data. The ratios of posterior probabilities, the *posterior odds*, of two models then directly follow as

$$\frac{\Pr(H_i|\mathbf{D})}{\Pr(H_j|\mathbf{D})} = \frac{\Pr(\mathbf{D}|H_i)}{\Pr(\mathbf{D}|H_j)}\frac{\Pr(H_i)}{\Pr(H_j)}. \tag{6.7}$$

In words, the posterior odds is given by the *prior odds* $\Pr(H_i)/\Pr(H_j)$ multiplied by the *Bayes factor* $B_{ij} = \Pr(\mathbf{D}|H_i)/\Pr(\mathbf{D}|H_j)$, which quantifies how much better H_i describes that data than H_j. The prior odds quantifies how much more plausible one model is than the other a priori, i.e., without considering the data. If there is no reason to favor one of the models over the other, the prior odds should be taken to equal unity (in which case the posterior odds equals the Bayes factor), but sometimes one must consider this point more carefully.

If one works under the assumption that precisely one of the considered models is correct, then it follows that $\Pr(\mathbf{D}) = \sum_{j=1}^{r}\Pr(\mathbf{D}|H_j)\Pr(H_j)$, and hence the posterior probabilities can be calculated as

$$\Pr(H_i|\mathbf{D}) = \frac{\Pr(\mathbf{D}|H_i)\Pr(H_i)}{\sum_{j=1}^{r}\Pr(\mathbf{D}|H_j)\Pr(H_j)} = \frac{1}{1+\sum_{j\neq i}\frac{\mathcal{Z}_j}{\mathcal{Z}_i}\frac{\Pr(H_j)}{\Pr(H_i)}}. \tag{6.8}$$

If the model H is *simple*, i.e., has no free parameters, then $\Pr(\mathbf{D}|H)$ is simply the probability (density) of the data $\mathbf{D}$ when H is assumed to be true and is the *likelihood* of that model. However, a model often does not uniquely specify its predictions, which instead depend on a number of free parameters $\boldsymbol{\Theta}$ of the model. These can either be allowed to take on a discrete or continuous set of values, with the latter being more common. For continuous parameters, one makes the replacement $\Pr(\boldsymbol{\Theta}) \to \Pr(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta}$, where now Pr stands for a probability *density* and N is the number of parameters. In the usual manner, sums in equations such as Eq. (6.5) become integrals, and, since under the assumption that H is the correct model one of the hypotheses labeled by values of model parameters must be correct, one obtains that the *evidence* of the model H is given by

$$\begin{aligned}\mathcal{Z} \equiv \Pr(\mathbf{D}|H) &= \int \Pr(\mathbf{D},\boldsymbol{\Theta}|H)\mathrm{d}^N\boldsymbol{\Theta} = \int \Pr(\mathbf{D}|\boldsymbol{\Theta},H)\Pr(\boldsymbol{\Theta}|H)\mathrm{d}^N\boldsymbol{\Theta} \\ &= \int \mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta}.\end{aligned} \tag{6.9}$$

Here, the *likelihood function* $\Pr(\mathbf{D}|\boldsymbol{\Theta},H)$ is the probability (density) of the data as a function of the assumed free parameters, which we often denote by $\mathcal{L}(\boldsymbol{\Theta})$ for

simplicity. The quantity $\Pr(\boldsymbol{\Theta}|H)$ is the prior probability (density) of the parameters and is often denoted by $\pi(\boldsymbol{\Theta})$. The prior should reflect how plausible different values of the parameters are, assuming the model to be correct, but not taking into account the data. It should always be normalized, i.e., integrate to unity. The assignment of priors are probably the most discussed and controversial part of Bayesian inference. This assignment is often far from trivial, but a very important part of any Bayesian analysis. This point will be discussed further in Sec. 6.3. Since, according to Eq. (6.9), the evidence is the likelihood marginalized over all the parameters, it is often called the *marginal likelihood*, and since $\Pr(\mathbf{D}|H)$ is the probability of the observed data assuming the model H, it can also be called the *model likelihood*.

This method of comparing models presented here is usually called *model selection*, although *model comparison* or *model inference* might be more accurate descriptions in the case that no model is actually selected. This is often compared and contrasted with the Bayesian method of estimating parameters of a fixed model, which will be described later. Although the practical methods and techniques can be rather different in these two levels, they are both simply applications of the laws of probability to perform inference, and there is no real fundamental difference between them.

From Eq. (6.9), we note that the evidence is the average of the likelihood over the prior, and hence this method automatically implements a form of *Occam's razor*, since usually a "simpler"[4] theory with a smaller parameter space will have a larger evidence than a more complicated one, unless the latter can fit the data substantially better. Usually, the data $\mathbf{D}$ used in Bayes' theorem to obtain the posterior inference is fixed to the value actually observed in experiments, but one can also consider $\Pr(\mathbf{D}|H)$ as a function of $\mathbf{D}$. Then, $\Pr(\mathbf{D}|H)$ is the probability with which the model predicted $\mathbf{D}$, and it is normalized to unity as long as the likelihood is correctly normalized. Since it is this probability evaluated at the data that has been observed that enters into the model comparison, a model is judged on how well the model as a whole predicted the data that subsequently was observed. The different predictions for a simple and more complex model is illustrated in Fig. 6.1.

This is rather different from standard methods used in particle physics (see Sec. 6.6) which mostly focuses on whether a model prohibits, or is incompatible with, the observed data. As one could expect, this can lead to overly complex models being preferred over simpler ones, since these will by construction be *compatible* with a large variety of data. At the same time, these models make very vague predictions, and hence are disfavoured from a Bayesian viewpoint. The total probability of the predicted data is always unity, and a more complex model has to spread this probability more "thinly" over the data space. It is only if the observed data is sufficiently incompatible with all other simpler and more predictive models that the more complex one is favoured.

[4] See Ref. [161] for further discussion on the notion of "simple" models.

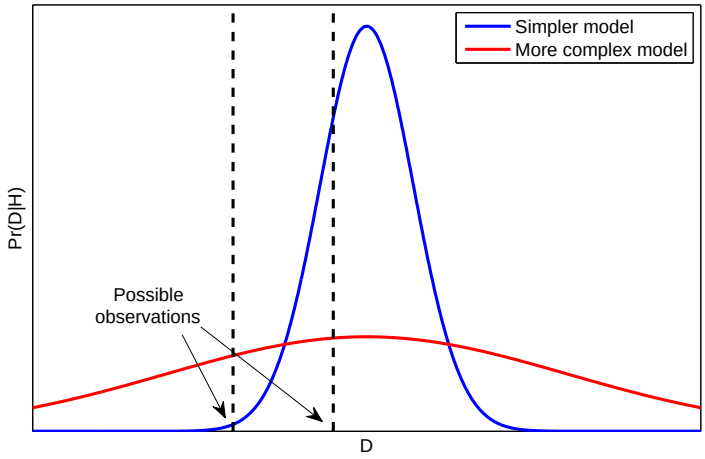

Figure 6.1. The predictions in data space of a simple and a more complex model. The Bayes factor of the two models is the ratio of the values of the respective probability distributions at the observed value of the data. Depending on which value is observed (exemplified by the dashed lines), either of the two models can be favoured by the data.

If a model is extended by additional parameters which are not constrained by the data, this will not lead to a smaller evidence, but instead leave the evidence unaffected. More generally, additional unconstrained parameter combinations are not disfavoured. This is an important point which is often misunderstood.[5]

The probabilities of the different hypotheses give the complete posterior inference on the space of models, and, as probabilities, these have a somewhat unique and meaningful interpretation on their own. In any case, Bayes factors, or rather posterior odds, are usually interpreted or "translated" into ordinary language using the so-called *Jeffreys scale*, given in Tab. 6.1 ("log" is the natural logarithm). This has been used in applications such as Refs. [166–170], although slightly more aggressive scales have been used previously [171, 172].

We also note that inference is never performed in complete isolation of any background information. Relevant such information could be details of the experimental setup, the structure of the theoretical models, the physical meaning of the free parameters, and previous measurements. All probabilities are always conditioned on this information, so that $\Pr(X|I)$ should really be written instead of $\Pr(X)$. For

[5] See, e.g., Ref. [163] (and rebuttal [164]) and Ref. [165] for particularly incorrect and misguided arguments against model comparison.

$\lvert\log(\text{odds})\rvert$	odds	$\Pr(H_1\lvert\mathbf{D})$	Interpretation
< 1.0	$\lesssim 3:1$	$\lesssim 0.75$	Inconclusive
1.0	$\simeq 3:1$	$\simeq 0.75$	Weak evidence
2.5	$\simeq 12:1$	$\simeq 0.92$	Moderate evidence
5.0	$\simeq 150:1$	$\simeq 0.993$	Strong evidence

Table 6.1. Jeffrey's scale often used for the interpretation of model odds. The posterior model probabilities for the preferred model H_1 are calculated by assuming only two competing hypotheses.

simplicity, we will use the shorter notation, but it is good to keep this conditioning in mind, since forgetting it can lead to apparent paradoxes when comparing probabilities with different background information [161].

Another advantage of the Bayesian approach is the possibility to take into account information which has not been obtained in the form of data. An important example is the case when the predictions of a model are uncertain, i.e., not possible to calculate exactly. These *theoretical* uncertainties can often by described as uncertainties in a set of additional parameters, and this uncertainty can then consistently be taken into account by averaging (integrating) over these parameters. In practice, this is done by adding these additional parameters to the parameter space, and using the external information on these parameters to assign priors to them. A relevant example is that of the nuclear matrix elements in Sec. 3.3, which have rather large theoretical uncertainties. That these uncertainties can be averaged over in a Bayesian analysis was exploited in paper V of this thesis [5].

The application in particle physics of Bayesian methods in general, and model comparison in particular, has been rather limited. Examples of the latter are analyses of supersymmetric models [169, 170], of real or simulated spectra in specific experiments [173–176], and of bottom quark decays [177].

6.2.1 Parameter inference

If one assumes a particular parmetrized model to be correct, the complete inference of the parameters of that model is given by the posterior distribution through Bayes' theorem

$$\Pr(\boldsymbol{\Theta}|\mathbf{D}, H) = \frac{\Pr(\mathbf{D}|\boldsymbol{\Theta}, H)\Pr(\boldsymbol{\Theta}|H)}{\Pr(\mathbf{D}|H)} = \frac{\mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{\mathcal{Z}}. \tag{6.10}$$

We see that the evidence, which appeared previously as a "likelihood" in the numerator, here appears as a normalization constant in the denominator. Since the evidence does not depend on the values of the parameters $\boldsymbol{\Theta}$, it is usually ignored in parameter estimation problems and the parameter inference is obtained using the unnormalized posterior. However, one can wonder how interesting the posterior of the parameters of a model assumed to be valid is if one has no information if this model is preferred by the data or not. The posterior distribution of the parameters

is only really relevant if the model does not have a very small posterior probability, since otherwise the model as a whole is strongly disfavored. In practice, this means that one should *first* calculate the evidences and posterior odds and only then, for the models with not to small evidences, calculate the posterior distribution.

In fact, if more than one model has a significant probability, it might be better to consider the distributions of parameters not assuming the model with maximum probability to be correct, but instead take into account the model uncertainty, giving the model-averaged posterior distribution [172, 178]

$$\Pr(\eta|\mathbf{D}) = \sum_{i=1}^{r} \Pr(\eta|H_i, \mathbf{D}) \Pr(H_i|\mathbf{D}), \tag{6.11}$$

which is the average of the individual distributions over the space of models, with weights equal to the posterior model probabilities. This can be done for both the prior and posterior distributions, but obviously the parameters η, which could be derived, need to be well-defined in all the models. The posterior in Eq. (6.10) is obtained by setting all prior model probabilities except one to zero.

Often, one wants to reduce the information contained in the full posterior in order to obtain the constraints on a subset of the parameters. From probability theory, we know that the probability density of any subset η of the parameters $\mathbf{\Theta} = (\eta, \rho)$ is obtained by integrating over the remaining parameters ρ, fully taking into account their uncertainties, as

$$\Pr(\eta|X) = \int \Pr(\eta, \rho|X) \mathrm{d}\rho. \tag{6.12}$$

This can be generalized to obtain the probability density of any (unique) function of the parameters $K = F(\mathbf{\Theta})$ as

$$\Pr(K|X) = \int \Pr(K|\mathbf{\Theta}, X) \Pr(\mathbf{\Theta}|X) \mathrm{d}^N\mathbf{\Theta} = \int \delta(K - F(\mathbf{\Theta})) \Pr(\mathbf{\Theta}|X) \mathrm{d}^N\mathbf{\Theta}. \tag{6.13}$$

Although this might look like a daunting integral, if one has access to samples from $\Pr(\mathbf{\Theta}|X)$, one can easily find the total probability in an interval of K by simply binning the samples.

The main result of Bayesian parameter inference is the posterior and its marginalized versions (usually in one or two dimensions). However, it is also common to give point estimates such as the posterior mode, mean, or median, as well as *credible intervals (regions)*, which are defined as intervals (regions) containing a certain amount of posterior probability. However, these regions are not unique without further restrictions, and in general they contain very little of the information that the full distribution contains.

Note that the posterior distributions of the parameters do not depend on the overall scale of the likelihood, since when rescaling the likelihood by a fixed constant, the evidence scales accordingly. The same holds true for the Bayes factors

and posterior model probabilities, since the ratio of evidences in Eq. (6.7) is also independent of the likelihood normalization.

6.3 Priors and sensitivity

The laws of probability dictate how the plausibilities of different propositions are related and can be manipulated; they provide a "grammar" for inference. However, in order to have something to manipulate, one has to give inputs by specifying or assigning probabilities. For normal inference, this is given by the probabilities $\Pr(\mathbf{D}|\mathbf{\Theta}, H)$, the "sampling distribution", as well the prior $\pi(\mathbf{\Theta})$, and potentially model priors as well.

Bayesian inference is subjective in the sense that the plausibility of different hypotheses depend on the prior information available. Furthermore, to represent this plausibility by a unique mathematical function can in many cases be difficult. Hence, instead of seeking a unique "right" prior for a given problem, one should evaluate how robust the posterior inference is when different sets of reasonable priors are assigned. Such a *sensitivity analysis* can then determine if the resulting inference depends severely on a particular choice of prior. In general, if the "correct" model (or a good approximation thereof) is in the considered set of models, one expects that model to be strongly preferred, independent of any (not unreasonable) prior assumptions, once enough data as been collected and analyzed. But how much is "enough"? An attempt to answer this question (which is difficult with standard methods) can be made by performing a sensitivity analysis and checking if the posterior inferences are robust to changes in the prior.

Of particular importance is that continuous prior and posterior probability densities, in order to keep the total probability invariant, transform under a change of variables $\mathbf{\Theta} \rightarrow \mathbf{\Omega} = \mathbf{\Omega}(\mathbf{\Theta})$ by multiplication by the Jacobian determinant, i.e., as

$$\Pr(\mathbf{\Omega}) = \Pr(\mathbf{\Theta}) \left| \frac{\partial \mathbf{\Theta}}{\partial \mathbf{\Omega}} \right|. \tag{6.14}$$

Hence, a prior uniform in one parameter will not be so in a nonlinear function of it, and thus the specification of a prior is essentially equivalent to the specification of a variable in which the prior is uniform.

In practice, the posterior of the parameters within a given model becomes insensitive to the details of their priors relatively fast, while the evidence is well-known for being more sensitive to the choice of prior distribution. In particular, if the prior is altered so that more or less prior probability is put in the region of negligible likelihood, but the details of the prior in the high-likelihood region remains the same, the posterior distribution will be practically unchanged, while the evidence, being the average of the likelihood under the prior, can change substantially. This is the reason why sometimes choosing a very broad (or an improper, as an extreme case) prior can sometimes lead to complete failure. This can create difficulties when one considers abstract problems without any connections to the real world, but in real

physics problems one has additional information in the form of the properties of the physical models one is examining, making it possible to assign meaningful and sensible priors in most cases. Evaluating the dependence of Bayes factors on the priors of the parameters is hence even more important than checking the sensitivity of the posterior distributions.[6]

6.3.1 Symmetries

Although, most of the time, assigning a prior probability distribution basically amounts to use the information at hand to simply pick a "reasonable" distribution, in some cases one can use mathematical and theoretical arguments to constrain the form of the prior [161, 180–182].

One way to do this is to look at the structure of the parameter space itself and find relevant symmetries which can constrain the form of the prior. The simplest case would be if one would have n discrete possible values for a quantity, and no information with which one could distinguish between them. Then, the problem will be invariant under permutations of these possibilities, and the only probability assignment consistent with this invariance would be the uniform one, $p_i = 1/n$ for each possibility.

The above argument can be generalized to continuous parameters. If the information we have of a particular problem is invariant under shifts $\mu \to \mu + a$ of a parameter μ for all $a \in \mathbb{R}$, then the only distribution consistent with this invariance is the uniform prior in μ, while for a *scale parameter* $\sigma > 0$, invariance under $\sigma \to \rho\sigma$ for all $\rho > 0$, implies a prior uniform in $\log \sigma$. These priors can be identified as corresponding to the *Haar measures* on the groups $(\mathbb{R}, +)$ and $(\mathbb{R}^+, \cdot)$, respectively. The problem with these priors is that they are *improper*, i.e., not normalizable, which can be traced back to the fact that the problem was assumed to be invariant under *arbitrary* transformations of the parameters. In most realistic problems, such invariances are not expected to be exact, since we usually have at least some vague idea of which parameter values are most plausible. However, if the symmetry is a good approximation in some region of parameter space, one can still use the approximate symmetry to motivate a certain form for the prior in that region.

These arguments can be generalized to any case where the parameter space can be identified with a group, in which case the Haar measure on that group is defined as the measure which is invariant under the action of all group elements, i.e., $\mu(aS) = \mu(S)$ (left Haar measure) or $\mu(Sa) = \mu(S)$ (right Haar measure) for subsets $S \subset G$ and elements $a \in G$. If G is either compact or commutative, the left and right Haar measures coincide.

These arguments actually have relevance to neutrino physics, something which is worth discussing in some more detail. Since the set of all possible neutrino mass

[6] One can either perform this for a small set of prior choices, or even better, to plot the Bayes factor as a function of the additional (hyper-)parameters specifying the prior as in, e.g., Ref. [179].

matrices (Dirac or Majorana) do not form a group, they do not have Haar measures. However, the mass matrix $\mathcal{M}_\nu$ is always diagonalized by unitary transformations as in Eq. (2.48) (or similar to Eq. (2.41) for Dirac neutrinos). One then demands that the measure over the neutrino mass matrix elements must not depend on which basis the matrix is defined in, i.e., for Majorana neutrinos one demands invariance of the measure under $\mathcal{M}_\nu \to V\mathcal{M}_\nu V^T$,[7] for $V \in \mathrm{U}(3)$ (or SO(3) for real matrices).[8] This is achieved if the measure over $\mathcal{M}_\nu$ contains the measure over U which is invariant under unitary or orthogonal rotations $U \to VU$, i.e., it contains the left Haar measure on U(3) (or SO(3)) (which equal the right due to compactness). The total measure on $\mathcal{M}_\nu$ is still not determined, but it can be written as a product of the measure over the neutrino masses and the Haar measure of the mixing matrix. In a sense, the 6-dimensional symmetry of the neutrino mass matrix has fixed the measure on six of the independent parameters (the mixing angles and phases) and there remains only an invariant but otherwise arbitrary measure on the eigenvalues.

However, if one in addition demands that all the elements of the mass matrix should be independent, these remaining degrees of freedom disappear, and the allowed measure becomes unique. It becomes a Gaussian [185]

$$\pi(\mathcal{M}_\nu) \propto \exp\left(-\frac{\mathrm{Tr}(\mathcal{M}_\nu \mathcal{M}_\nu^\dagger)}{2m_0^2}\right) \tag{6.15}$$

for some arbitrary mass scale m_0. This can be seen as a generalization of the so-called Herschel-Maxwell derivation of the Gaussian distribution for the case of a U(1) symmetry: If two quantities x and y are independent and their distribution only depends on the "radius" $\sqrt{x^2+y^2}$, then that distribution must be a Gaussian with some undetermined width.

Although the above arguments leading to a unique measure on the neutrino mixing parameters have been known for more than a decade [183], it has not been realized that the use of Haar measures is a well-known technique for constraining prior distributions in the statistics literature. Hence, the proper meaning of such a measure on the neutrino mass matrix as a Bayesian prior probability distribution have not been pointed out. Instead, these distributions were mostly though of as frequentist sampling distributions in some way.[9] However, such an interpretation makes no sense since the mixing parameters, as all other parameters, have fixed values (assuming the model is correct), and are not random (see also Sec. 6.1).[10]

[7] For the case of Dirac neturinos and further discussion, see Refs. [183,184] for further discussion.
[8] This can of course be generalized to an arbitrary number n of neutrino species.
[9] See, for example, Refs. [183, 184, 186, 187].
[10] One could conjecture that we live in one "sample" out a large number of universes, which all have different physical parameters generated according to some distribution, but the testability of such a hypothesis is rather questionable.

6.3.2 Maximum entropy

Sometimes there is information available that is relevant and one would like to use to constrain the prior distribution, but this information is not in the form of data, and hence cannot be used to construct a likelihood. As an example, say we want to assign probabilities to the six possible outcomes of the next roll of a die, and we are told nothing but the fact that the average outcome in a large number of rolls is, say, 4. This information is *testable* in the sense that for any given probability distribution over the different outcomes, one can always determine if the distribution is consistent with the information. However, there might be many distributions which are consistent with the given information, so the question is which one should be preferred. In our example, assigning probability 1 to obtaining a 4 in the next roll (and zero to all other outcomes) is consistent with the information, but is rather unreasonable since it excludes possible outcomes without any reason.

The assignment which should be made is the one which is consistent with the given information but imposes no other, which is the distribution which contains the least amount of information or the most *entropy* [161, 162, 181], which for n discrete possibilities is given by

$$\mathcal{H} = -\sum_{i=1}^{n} p_i \log p_i. \tag{6.16}$$

For a continuous variable, the naive replacement of the sum by an integral is problematic, so one must introduce an additional density $m(\boldsymbol{\Theta})$, which can be taken as the probability distribution before taking into account the additional information, and use the entropy

$$\mathcal{H} = -\int p(\boldsymbol{\Theta}) \log\left(\frac{p(\boldsymbol{\Theta})}{m(\boldsymbol{\Theta})}\right) \mathrm{d}^N\boldsymbol{\Theta}. \tag{6.17}$$

The negative of this quantity is commonly known as the *Kullback-Leibler divergence* between the distributions m and p and is a measure of the information gained when updating m to p. If data is used to update the prior to the posterior distribution, the Kullback-Leibler divergence between these measures the information contained in the data, which can be interesting on its own.

Often, the information one wants to consider is in the form of constraints on n expectations $G_i = \int p(\boldsymbol{\Theta}) g_i(\boldsymbol{\Theta}) \mathrm{d}^N\boldsymbol{\Theta}$, to which the solution can be found using variational methods as $p(\boldsymbol{\Theta}) = m(\boldsymbol{\Theta}) \exp\left(\sum \lambda_i g_i(\boldsymbol{\Theta})\right) / Z(\lambda_1, \cdots, \lambda_n)$, with Z and all the λ's to be determined using the constraints. If there are no constraints in addition to normalization, $p(\boldsymbol{\Theta}) = m(\boldsymbol{\Theta})$.

Important special cases include those with a broad, roughly uniform m. For a positive quantity with a fixed mean, the maximum entropy distribution is an exponential distribution. For a real-valued quantity with fixed mean and variance, a Gaussian distribution is obtained. However, not all classes of distributions contain

a maximum entropy distribution, such as the distributions on the real numbers with only a specified mean.

6.4 Combining and comparing data

When analyzing experimental results one often wants to combine data from different sources in order to obtain the total constraints on the models and their parameters. Often, the different data sets can be assumed to be statistically independent once a complete set of parameters that specify the individual distributions are given. If the full set of data is given by $\mathbf{D} = (D_1, D_2, \ldots, D_n)$, the likelihood becomes

$$\mathcal{L}(\mathbf{\Theta}) = \Pr(\mathbf{D}|\mathbf{\Theta}, H) = \prod_{i=1}^{n} \Pr(\mathbf{D}_i|\mathbf{\Theta}, H) = \prod_{i=1}^{n} \mathcal{L}_i(\mathbf{\Theta}). \tag{6.18}$$

It is important to remember that the individual distributions multiply only when they are conditioned on all the parameters needed to fully specify all the distributions. If some of the parameters are marginalized over or eliminated in some way, the likelihoods *do not* generally multiply. In particular, for the evidence when using two independent data sets, $\Pr(D_1, D_2|H) \neq \Pr(D_1|H)\Pr(D_2|H)$, and so $\Pr(D_1|D_2, H)$ depends on D_2 in general.

Now, the usual path of inference would use the combination of data to infer which models can best describe the data, and which values the parameters of those models are most likely to have. It can also be useful to perform some sort of *checks* of the individual models, i.e., evaluating the general compatibility of the data with the model at hand without performing rigorous inference by comparing it with other models. This will be discussed further in Sec. 6.6

Another important method is to check if the parameter constraints from different data sets are consistent within a particular model. In fact, this can be thought of as checking if the different data sets can be described by a single (unknown) value of the parameters, or if they require different values. In a Bayesian context, this can be formulated as a model comparison problem [188], originally applied to cosmological models. In particle physics, it has been applied in [5, 169, 189].

We have a full set of data given by $\mathbf{D} = (D_{\text{test}}, D_{\text{bkg}})$, where we want to test if $D_{\text{test}} = (D_1, D_2, \ldots, D_n)$, with $n \geq 2$, gives consistent constraints on the parameters of a model H. The quantity D_{bkg} is a set of possible "background data" that we do not want to check the consistency of, but take as given. It can be thought of as being included in the background information and acts like a prior constraint on the model parameters. Then consider the different assumptions

C: The data D_{test} all give consistent parameter constraints (within H), given the background data D_{bkg}. Hence, all the data can be described by a single set of parameters (of H).

$\overline{C}$: The data D_{test} are inconsistent and hence lead to different regions of parameter space being preferred, i.e., $(D_1, D_2, \ldots, D_n)$ need different sets of parameters to describe the data.

The posterior odds of C and $\overline{C}$ (with implicit conditioning on H) is then given by

$$\frac{\Pr(C|D_{\text{test}}, D_{\text{bkg}})}{\Pr(\overline{C}|D_{\text{test}}, D_{\text{bkg}})} = \frac{\Pr(D_{\text{test}}|D_{\text{bkg}}, C)}{\Pr(D_{\text{test}}|D_{\text{bkg}}, \overline{C})} \frac{\Pr(C|D_{\text{bkg}})}{\Pr(\overline{C}|D_{\text{bkg}})} = \frac{\Pr(D_{\text{test}}|D_{\text{bkg}}, C)}{\Pr(D_{\text{test}}|D_{\text{bkg}}, \overline{C})} \frac{\Pr(C)}{\Pr(\overline{C})}, \tag{6.19}$$

where the last step follows from the fact that $\Pr(C|D_{\text{bkg}})/\Pr(\overline{C}|D_{\text{bkg}})$ equals $\Pr(C)/\Pr(\overline{C})$, since the probability that D_{test} is consistent should not change without considering it. From this also follows that $\Pr(D_{\text{bkg}}|\overline{C}) = \Pr(D_{\text{bkg}}|C) = \Pr(D_{\text{bkg}})$. The calculable part of Eq. (6.19) is the Bayes factor

$$\mathcal{R} = \frac{\Pr(D_{\text{test}}|D_{\text{bkg}}, C)}{\Pr(D_{\text{test}}|D_{\text{bkg}}, \overline{C})} = \frac{\Pr(D_{\text{test}}|D_{\text{bkg}})}{\prod_{i=1}^{n} \Pr(D_i|D_{\text{bkg}})}, \tag{6.20}$$

where the last step follows from the way that the hypotheses C and $\overline{C}$ were defined, i.e., the data in D_{test} can be described by the same parameters under C, but need different sets under $\overline{C}$. These evidences are given by the integrals of the likelihood over the "prior" $\Pr(\boldsymbol{\Theta}|D_{\text{bkg}})$ as in

$$\Pr(D_{\text{test}}|D_{\text{bkg}}) = \int \Pr(D_{\text{test}}|\boldsymbol{\Theta}) \Pr(\boldsymbol{\Theta}|D_{\text{bkg}}) \mathrm{d}^N\boldsymbol{\Theta}, \tag{6.21}$$

and similarly for the other evidences. The conditioning on D_{bkg} can be dropped in the likelihood since the probability distribution of the data D_{test} does not depend on D_{bkg} if all the free parameters $\boldsymbol{\Theta}$ are fixed. If there is no background data, one of course simply uses the "original" priors in the evidence integrals.

These integrals can however be difficult to perform unless $\Pr(\boldsymbol{\Theta}|D_{\text{bkg}})$ has a simple form. In this case one can use that Eq. (6.20) equals

$$\mathcal{R} = \frac{\Pr(D_{\text{test}}, D_{\text{bkg}})}{\prod_{i=1}^{n} \Pr(D_i, D_{\text{bkg}})} \Pr(D_{\text{bkg}})^{n-1}. \tag{6.22}$$

These evidences are the ones evaluated using the original priors, but now also including the background data in the likelihood.

After having derived the expressions for the Bayes factor in Eqs. (6.20) and (6.22) it is useful see what happens when it is applied on simple problems where it is known what the "correct" result should be. This has been done on both simple and more advanced toy problems in Refs. [169, 188, 189] for the case $n = 2$, i.e., considering only two data sets. Some simple analytical limiting cases can be constructed by letting D_1 and D_2 result in likelihoods $\mathcal{L}_1(\boldsymbol{\Theta})$ and $\mathcal{L}_2(\boldsymbol{\Theta})$, respectively. If one of the likelihoods (say $\mathcal{L}_2$) is constant, and hence give no information on the parameter values, one should obtain $\mathcal{R} = 1$, since in this case there is only one actual measurement and hence one cannot say anything about the consistency of

the data sets. This is indeed what Eq. (6.22) reduces to. As a second example, consider the case when the second data set determines the model parameters exactly, $\mathcal{L}_2(\boldsymbol{\Theta}) = \delta(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)$. In this case the consistency test should be equivalent to testing if the first data set favors $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ or the more complex model with prior $\pi(\boldsymbol{\Theta})$ (or $\Pr(\boldsymbol{\Theta}|D_{\rm bkg})$ if background data is included). Indeed, one finds that $\mathcal{R} = \Pr(D_1|\boldsymbol{\Theta}_0)/\Pr(D_1|H, D_{\rm bkg}) = \mathcal{L}_1(\boldsymbol{\Theta}_0)/\int \mathcal{L}_1(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta}|D_{\rm bkg})\mathrm{d}^N\boldsymbol{\Theta}$.

Finally, we note that one can also combine the data sets into a single likelihood, but keep open the possibility of an error in the modelling of the data, such as biases or incorrectly estimated uncertainties. As a general approach not requiring any detailed knowledge of what might be wrong with the data sets [190], one can introduce a set of additional parameters $\alpha_1, \ldots, \alpha_n$ in the form of weights of the individual likelihoods, so that the factors $\Pr(\mathbf{D}_i|\boldsymbol{\Theta}, H)$ in the likelihood are replaced by

$$\Pr(\mathbf{D}_i|\boldsymbol{\Theta}, \alpha_i, H) = \frac{\Pr(\mathbf{D}_i|\boldsymbol{\Theta}, \alpha_i = 1, H)^{\alpha_i}}{Z_i(\boldsymbol{\Theta}, \alpha_i)}, \tag{6.23}$$

where $Z_i(\boldsymbol{\Theta}, \alpha_i)$ are required normalization factors. Hence, for $\alpha_i = 1$ the data set is included in the usual way and for $\alpha_i = 0$ it is completely excluded. One can then let the data determine whether the inclusion of these weights are necessary by comparing the model where all weights equal one (the standard case) to the model where the weights are free (through the Bayesian evidence, as usual). If the data sets are sufficiently discordant, the model with nonunitary weights will be preferred. Once the full posterior has been marginalized down to the physical parameters, it will in this case often be multimodal, with each mode corresponding to the region preferred by one of the data sets. If the sets are compatible, the model with unitary weights will be favored and one will obtain the regular inference, without the risk of having increased uncertainty on the parameter values originating from the additional freedom in the extra parameters.

6.5 Numerical methods and approximations

To calculate the posterior distribution, and in particular the evidence, is not in general an easy task. In realistic problems, the parameter spaces often have high dimensionalities and the likelihoods are often significantly non-zero only in a very small region of the parameter space. Furthermore, likelihoods can often be quite complicated functions, having multiple modes or strong degeneracies, confining the posterior to thin sheets in the parameter space.

The simplest approach to obtaining the evidence and the posterior is to use a grid in the parameter space on which one evaluates the likelihood and the prior. However, this become impractical even for problems with quite a small number of dimensions since the number of points required to produce a dense enough grid grows exponentially with the dimensionality. This could be improved upon using deterministic but adaptive methods, but also those will fail for all but the simplest problems.

To overcome this *curse of dimensionality*, the usual method is to use *Monte Carlo methods* which rely on using random numbers to obtain the desired numerical results. For parameter estimation, the standard choice has long been *Markov-Chain Monte Carlo* (MCMC), in particular in the form of the *Metropolis-Hastings* algorithm [191, 192], to obtain a set of samples from the posterior. Although it usually works quite well in cases where the likelihood is rather "well-behaved", it can have difficulties when the likelihood is not so simple. In addition, it does not give a straightforward estimate of the evidence.

To calculate the evidence, the most naive Monte Carlo approach is to simply draw randomly from the prior distribution and then calculate the average of the likelihoods of those samples. However, if there is only a small volume of the parameter space in which the likelihood is large, none or a very small number of the sampled points will fall into the high-likelihood region, leading to a very bad evidence estimate. If one has access to samples from the posterior, one could use the harmonic mean of the likelihoods of those samples to form an estimate [172], since $\mathcal{Z}^{-1} = \int \mathcal{Z}^{-1}\pi(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta} = \int \mathcal{L}(\boldsymbol{\Theta})^{-1}P(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta}$. However, this method is notoriously unreliable since the estimate becomes dominated by rare low-likelihood points in the tails of the distribution, and the estimate generally has very large (or even infinite) variance.

In order to evaluate the evidence one must instead in some way utilize a distribution in between the two extremes above (i.e., the prior and the posterior), so that one can cover the whole of parameter space, but at the same time get adequate exploration of the peak of the posterior. The canonical way to do this, called *thermodynamic integration*, is to obtain MCMC samples from a range of distributions proportional to $\mathcal{L}(\boldsymbol{\Theta})^\lambda\pi(\boldsymbol{\Theta})$, for different values of the *inverse temperature* λ. Defining $\mathcal{Z}(\lambda) \equiv \int \mathcal{L}(\boldsymbol{\Theta})^\lambda\pi(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta}$, the evidence evaluation reduces to a one-dimensional integral

$$\log \mathcal{Z} = \int_0^1 \frac{\mathrm{d}\log \mathcal{Z}(\lambda)}{\mathrm{d}\lambda}\mathrm{d}\lambda = \int_0^1 \langle \log \mathcal{L} \rangle_\lambda \, \mathrm{d}\lambda, \tag{6.24}$$

where $\langle\cdot\rangle_\lambda$ denotes the average over the posterior at inverse temperature λ, i.e., over the distribution proportional to $\mathcal{L}(\boldsymbol{\Theta})^\lambda\pi(\boldsymbol{\Theta})$. This average can be calculated using standard methods, such as MCMC, but it still has the previously discussed difficulties for complex likelihood functions.

Nested sampling [193, 194] is another method which relies on sampling over a range of *nested* distributions proportional to $\pi(\boldsymbol{\Theta})\mathrm{I}(\mathcal{L}(\boldsymbol{\Theta}) > \mathcal{L}')$, where I is the indicator function and $\mathcal{L}'$ starts of at zero and increases until it is close to the maximum of the likelihood. Define the remaining prior volume as a function of the value of the likelihood as

$$X(\mathcal{L}') = \int_{\mathcal{L}(\boldsymbol{\Theta})>\mathcal{L}'} \pi(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta} \tag{6.25}$$

and let the inverse of this function be $\mathcal{L}(X)$. Then the evidence can be written as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)\mathrm{d}X. \tag{6.26}$$

This one-dimensional integral can be computed as long as $\mathcal{L}(X)$ can be found, which is accomplished by first drawing N_{live} *live points* from the prior $\pi(\boldsymbol{\Theta})$. At each subsequent iteration i, the point with the smallest likelihood, $\mathcal{L}_i$, is discarded and replaced by a new point drawn from the prior subject to the constraint that the likelihood exceeds $\mathcal{L}_i$, i.e., from the distribution $\pi(\boldsymbol{\Theta})\mathrm{I}(\mathcal{L}(\boldsymbol{\Theta}) > \mathcal{L}_i)/X(\mathcal{L}_i)$. This process, generating likelihoods $\mathcal{L}_i = \mathcal{L}(X_i)$, is then continued until the peak of the likelihood is reached, and the $\mathcal{L}_i$'s are then used to estimate the integral in Eq. (6.26). The dominating uncertainty of the evidence estimate comes not from the discretization or truncation of the integral, but instead from the fact that the X_i's are not known precisely. However, since their distributions are known, this uncertainty can be folded into the uncertainty of the evidence.

The main difficulty with constructing an efficient and general nested sampling algorithm is to enable efficient sampling from the constrained distribution, and to be able to do this for as general likelihood functions as possible. The most widely used such algorithm is MultiNest [195,196] which accomplishes this by utilizing the set of live point to perform an estimate of the iso-likelihood surface as a set of possibly overlapping ellipsoids, from which sampling is straightforward. The number of ellipsoids used is optimized, allowing for them to follow strong curving degeneracies and isolate many independent modes, while at the same time only requiring a small number of ellipsoids for simple problems. Compared to thermodynamic integration it is in practical applications often more efficient (i.e., requiring fewer likelihood evaluations). This is especially true for difficult problems, where even estimating the posterior, a generally easier task, can be very difficult with other methods. Also, thermodynamic integration will fail completely if $\log \mathcal{L}$ is not a concave function of $\log X$, which can happen in realistic problems [168, 193, 194].

Another approach for evaluating integrals (and generating samples from a distribution) is *importance sampling*, which is based on that for any distribution $Q(\boldsymbol{\Theta})$

$$\int \mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta} = \int \frac{\mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})} Q(\boldsymbol{\Theta})\mathrm{d}^N\boldsymbol{\Theta}. \tag{6.27}$$

Hence, the evidence can be accurately estimated if one can find a distribution Q for which samples can be obtained, $\pi(\boldsymbol{\Theta})/Q(\boldsymbol{\Theta})$ is calculable, and Q has "fat" tails enabling the whole prior volume to be sampled, but at the same time is highly peaked so that the full high-likelihood region is adequately explored. Now, for an algorithm based on importance sampling to be generally useful, allowing for its widespread use, the user should not need to specify a Q, but the algorithm should be able to automatically determine a suitable Q given the likelihood function. One way to accomplish this would be to use for Q a *mixture density* $Q(\boldsymbol{\Theta}) = M^{-1}\sum_{i=1}^{M} f_i(\boldsymbol{\Theta})/Z_i$, where the Z_i's are individual normalization constants and

each subsequent component f_i is chosen to be more concentrated around the peak of the likelihood then the previous ones. For example, both $f_i(\boldsymbol{\Theta}) = \mathcal{L}(\boldsymbol{\Theta})^{\lambda_i}\pi(\boldsymbol{\Theta})$ and $f_i(\boldsymbol{\Theta}) = \pi(\boldsymbol{\Theta})\mathrm{I}(\mathcal{L}(\boldsymbol{\Theta}) > \mathcal{L}_i)$ could be used, but this would require normalization constants $\mathcal{Z}(\lambda_i)$ and $X(\mathcal{L}_i)$, respectively, which are not known but would have to be estimated in the process.

Recently, the MultiNest algorithm was extended to perform importance sampling in parallel to nested sampling by using for each mixture component f_i a uniform density within the ellipsoidal decomposition at each iteration [197], for which the normalization constants are basically known. This allows for taking into account the information content present in all the points discarded in the nested sampling estimate, and is seen to give even more accurate evidence estimates in realistic problems. A slight complication, however, is that the ellipsoidal decompositions, as well as the number of samples taken from each such decomposition, are random variables too and depend not only on the likelihood but also on the specific set of points previously sampled (hence yielding dependent samples). This leads to an additional (but ideally negligible) contribution to the variance of the evidence estimate. See Ref. [197, 198] (and references therein) for further discussion on the similarities and differences between different methods of evidence evaluation.

We also note that with most methods, once the evidence has been estimated, samples from posterior distribution follow with no or very little additional work.[11]

The Bayes factor between a model H_1 and its restriction H_0 specified by fixing a subset of the parameters $\eta = \eta_0$ is (under some weak assumptions on the form of the prior) possible to calculate using only the marginalized posterior of η in H_1. It is given by the *Savage-Dickey density ratio*

$$\frac{\Pr(\mathbf{D}|H_0)}{\Pr(\mathbf{D}|H_1)} = \frac{\Pr(\eta_0|\mathbf{D}, H_1)}{\Pr(\eta_0|H_1)}, \tag{6.28}$$

and so the data favours $\eta = \eta_0$ if the data increases the density under the extended model at that point (and disfavors $\eta = \eta_0$ if the density decreases), which makes sense. The ratio is also manifestly invariant under arbitrary transformations of η. Also, this once again points out that one needs both the prior as well as posterior distribution to compare such nested models, and that the absolute values of the posterior density, as well as the ratios of the posterior at different values of η (such as η_0 and the maximum) are irrelevant. This representation can be useful since it only requires normalization of a low-dimensional distribution (that of η), regardless of the total dimensionalities of the parameter spaces, which can often be done numerically. A limitation is that it only works when comparing nested models and that numerical estimates can be quite noisy if there are few samples near η_0. See, e.g., Ref. [199] for further discussion of the Savage-Dickey density ratio and the related measure-theoretic paradox.

[11] If one has calculated the normalization constant of the posterior distribution, it is plausible that one has enough information to be able to obtain samples from it.

The evidence can be obtained analytically if one approximates $\mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$ by a Gaussian. Expand $\log(\mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta}))$ around the maximum $\hat{\boldsymbol{\Theta}}$, and denote the covariance matrix (inverse of the negative Hessian) by Σ. Then

$$\mathcal{Z} \simeq (2\pi)^{N/2} \det(\Sigma)^{1/2} \mathcal{L}(\hat{\boldsymbol{\Theta}})\pi(\hat{\boldsymbol{\Theta}}). \tag{6.29}$$

Here one can also see the compensating effects of higher likelihoods giving a higher evidence and increased "surprise" or "fine-tuning" entering through the covariance matrix and prior. The main limitation of this approximation is, obviously, that it only works for posteriors which are close to Gaussian. Although this might sometimes be a good approximation, often it will not.

There are also various information criteria, motivated either by Bayesian or information theory, designed to take into account goodness of fit and model complexity when comparing models [200]. In particular, the Bayesian Information Criterion (BIC) is an asymptotic approximation of the evidence, which is generally rather easily computed from only the maximum likelihood, and also does not explicitly depend on any prior. However, when the data is only moderately constraining (which is the case we are interested in, since otherwise we do not need any statistics in the first place) they do not generally approximate the evidence well at all, and so the evidence remains the preferred quantity for comparing models.

6.6 Frequentist methods

As noted previously, frequentist probability describes "randomness" and is based on a more restrictive notion as a relative frequency of an event in a large number of repeated trials. This interpretation, and the statistical methods associated with it, is dominating practical data analysis in the field of particle physics. Since we have already studied Bayesian methods in some detail, in this section I will make frequent comparisons with the Bayesian methods we have encountered so far in order to highlight their rather significant differences. Also, I will not be going into much details about the different methods, but rather try to describe the main ideas behind them.

In the frequentist approach, one is not allowed to speak about probabilities of hypotheses or propositions, so one must choose a different approach of evaluating different models. Since one still has probabilities of the data assuming a certain model with any free parameters fixed to be correct, there is no problem with calculating the likelihood, $\Pr(\mathbf{D}|H)$. However, this probability by itself says nothing about the validity of the model, and that probability can be arbitrarily small (even dimensionfull) even if the observed data is a typical model prediction, or even the most probable outcome. Relative probabilities $\Pr(\mathbf{D}|H_1)/\Pr(\mathbf{D}|H_2)$ of simple hypotheses are special cases of Bayes factors also defined in a frequentist sense, but it cannot be converted into posterior odds or probabilities. In principle, one could attempt to compare models using only the value of the likelihood ratio, but the

question is then what one should do if the models have free parameters, and on what scale the likelihood ratio should be interpreted.

The main idea behind the frequentist approach is instead to compare the data that was observed with data which *could* have been observed or what is expected to be observed if one would repeat the experiment many times. If one constructs some function of the data, called a *test statistic*, $T = T(\mathbf{D})$, and assume the truth of some hypothesis, T will take on different values under repetitions of an identical experiment, and hence it has a probability distribution in a frequentist sense. The hypothesis is then to be assessed by comparing the observed value of the statistic with the long-run distribution expected in hypothetical repetitions of the experiment. Although the choice of statistic and method of comparison is not unique, there are a few classes of statistics which are commonly used and known to often perform good in practice. One of them, the profile likelihood, will be discussed later. The remaining question is then how to compare the observed value of T with its expected distribution, and how to use this comparison to asses the model in question.

6.6.1 Hypothesis tests

When testing statistical models, there are two rather different approaches which in practice tend to get somewhat mixed up. The first is often called *hypothesis testing* (and often associated with J. Neyman and E. Pearson) and proposes that one should test a *null hypothesis* H_0 not in isolation, but against an alternative H_1 (similarly to Bayesian methods).

Construct a statistic T so that large values of T correspond to data which are "unexpected" under H_0. The test then consists of a decision problem: *reject* H_0 if $T \geq c$, where c is a *pre-specified* number. Then, one calculates and reports the *type-1* and *type-2* error probabilities $\alpha \equiv \Pr(\text{reject } H_0|H_0)$ and $\beta \equiv \Pr(\text{accept } H_0|H_1)$. The justification for this method is that if the same procedure is repeated on a large number of identical experiments, the fraction of times an error is made is given by α and β, respectively. One of its main drawbacks is that one can only report a pre-determined α regardless of the actual observed data. This necessarily discards almost all the information contained in the data and makes it difficult to figure out what the data actually implies (apart from the hypothesis being rejected or not). In principle, one could calculate the Bayes factor using only "H_0 is rejected" as data, which would be $\alpha/(1-\beta)$. However, this is not very useful since the "real" Bayes factor using all the data is generally very different. Another complication is that usually the alternative hypothesis (as well as the null) has free parameters, in which case β as function of those parameters must be determined, but reporting this can be impractical.

Due to the above drawbacks, this method is rarely applied in particle physics when testing pre-determined null hypotheses. Using a pre-specified α is, however, common when deriving upper limits on the strength of a possible signal and when constructing *confidence intervals* for a free parameter. In these cases, it is instead

the limits and intervals which vary with the data. When estimating parameters of a fixed model the most common method is to give the values of the parameters which maximize the likelihood as a point estimate. In order to estimate the uncertainty of this estimate, one can try to estimate its variance, or more commonly to calculate confidence intervals. These intervals are constructed in order to include the true value of the parameter a fraction $1 - \alpha$ of repeated experiments, and hence can be constructed by including all values which are not rejected at the level α. Since the construction of confidence intervals is not unique[12] there can often be large differences in the resulting confidence intervals. Near physical boundaries of the parameter space (where much of particle physics takes place) this can even result in no physical values of the parameters being included in the confidence interval due to downward fluctuations of the background. Although one can choose test statistics to prevent this from happening [201, 202], it remains that one can obtain very different results by choosing a slightly different (but still reasonable) method. Hence, although the intervals are constructed to include the true value in a large fraction of an ensemble of repeated experiments, a specific interval using the data that was actually observed can contain essentially no information about the underlying models. In these situations it is extremely important for the experimental collaborations to report enough additional information allowing others to as accurately as possible recover the information contained in the data. This is, however, unfortunately not always done.

6.6.2 P-values

The most common test when looking for new physical processes and testing prespecified hypotheses is instead *significance testing* (often associated with R. Fisher). The idea is that one should not be forced to compare a null hypothesis H_0 against a specific alternative, but one should be able to test the compatibility with the data of H_0 alone.

Constructing a test statistic T as above, one would calculate the *p-value* defined as a *tail probability* $p = \Pr(T \geq T_{\text{obs}}|H_0)$, which is the probability of obtaining a test statistic equal or larger than the value calculated for the observed data, T_{obs}, in a large number of repeated experiments, assuming H_0. The idea is then that the p-value would indicate the "strength of evidence" against H_0. A small p-value would lead us to doubt the validity of H_0, but a large one would not give evidence in favour of H_0. Compared to Neyman-Pearson hypothesis testing, this has the clear advantage that the reported strength of evidence actually varies with the data (similar to Bayesian methods), but it also has several drawbacks.

[12] As mentioned earlier, Bayesian credible intervals are not unique either, but these are far from the main output of the inference, and rather derived quantities containing very little of the information contained in the full posterior. Many particle physicists feel for some reason that an experiment *must* be completely summarized by a few numbers such as an interval, upper limit, or best-fit point, and that these might even be the main results of an experiment.

The motivation for using the probability of data more extreme than the observed data to evaluate a hypothesis is not clear, and it is well-known that interpreting the p-value as a posterior probability or an error rate severely overestimates the evidence against H_0. This can be partly understood by the fact the one essentially replaces the data $\mathbf{D} = \mathbf{D}_{\rm obs}$ not with the value of the test statistic, $T = T_{\rm obs}$, but with $T \geq T_{\rm obs}$. Since in all the data sets where this is true are more extreme than the actual observed data, it is not strange that one finds overly strong evidence against H_0.[13] Another possibility is the case when a large number of tests are performed on a null hypothesis, but the number of cases where it could be false cannot be very many.[14] As a consequence, most of the null hypotheses which are tested with the result being a "small" p-value (say, $p \simeq 0.01$) are in practice actually correct. Of course, physicists have realized this by trial and error a long time ago, and so are demanding much smaller p-values before claiming a discovery. But how should we then interpret p-values? When is a p-value "small" and when is it "very small"? We have a well-defined quantity which does not depend on considering any alternative model nor needs any priors to be specified. Instead (or rather because of this), there is a large uncertainty in its interpretation. The current consensus, within particle physics at least, is that any p larger than 0.01 is not significant evidence, while $p = 10^{-4}$ *is* significant, and $p = 10^{-6}$ can be considered "strong" evidence somehow. In order to reflect this, it is common to translate p-values into a "number of σ's" by relating it to the standard normal distribution, $S = \phi^{-1}(1-p)$, with ϕ^{-1} being the inverse of the standard normal cumulative distribution function. The value $S \simeq 5$ ($p \simeq 10^{-7} - 10^{-6}$) has become the norm in particle physics for claiming a "discovery", meaning a definite rejection of the background-only hypothesis, while S less then roughly 2.5 does not constitute any real evidence.

Having a universal interpretation of the p-value for all situations does of course not seem reasonable from a Bayesian point of view, in which one should also consider the nature of the alternatives one is comparing with. This fact, in particular when a "small" p-value is found for a null hypothesis simultaneously as a probabilistic analysis shows that the null is actually preferred over the alternatives, is often called *Lindley's paradox*.[15] However, even for Bayesians it can still be useful to use p-values (and also Bayesian extensions exist [203]) for *model checking* (as opposed to full probabilistic inference). In particular, there is the possibility that the correct model is not in the set of considered ones, and in this case even the most preferred model can be a bad description of the data, giving a small p-value. In this case one should consider extending the scope of the considered models. Finally, we note that Jeffreys [171] has made a rather famous comment on using p-values for testing: "A

[13] In principle, one could attempt to calculate the posterior (but only when considering an alternative model, as always) $\Pr(H_0|T \geq T_{\rm obs}) \leq \Pr(H_0|T = T_{\rm obs})$ which would presumably also exaggerate the evidence against H_0.

[14] Consider only the numerous searches of new particles at the LHC, where each search has a very small probability of actually being sensitive to the signal of a new particle.

[15] See, for example, Ref. [167] for further discussion.

hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred."

6.6.3 Profile likelihood ratio

We have so far not discussed how to choose the test statistic. In practice, choosing a test statistic of which you know the distribution precisely is only possible in very simple cases. In more realistic cases, one is forced to use statistics whose distributions are only approximately known, leading to the test also only being approximate. A very commonly used statistic is based on the *profile likelihood ratio*. If one wants to test the hypothesis that the data is generated by a model with some free parameters ρ, one then considers an extension of that model with additional parameters η so that the original model is obtained for $\eta = \eta_0$.[16] Then one calculates

$$Q^2(\mathbf{D}) \equiv -2\log\frac{\sup_\rho \mathcal{L}(\eta_0,\rho)}{\sup_{\eta,\rho}\mathcal{L}(\eta,\rho)} = -2\log\frac{\mathcal{L}(\eta_0,\hat{\hat{\rho}}(\eta_0))}{\mathcal{L}(\hat{\eta},\hat{\rho})}, \tag{6.30}$$

where "sup" denotes the supremum, a single hat denotes the parameters which maximize the likelihood, and a double hat indicates the conditional maximum for fixed η_0. A good thing about this statistic is that large values of the test statistic (which should give evidence against the hypothesis) correspond to cases when there is a considered hypothesis which can fit the data substantially better than the model being tested.

Under the assumption that the hypothesis is correct, then in the large sample limit Q^2 (often denoted by $\Delta\chi^2$) will have a χ^2-distribution with number of degrees of freedom equal to the dimensionality of η. This result is known as *Wilks' theorem* [204]. However, there are a number of conditions which need to be satisfied for this to hold [205]. One could of course go through the mathematical details, but the main idea is that Q^2 becomes χ^2-distributed when the distribution of the maximum likelihood estimates, $(\hat{\eta},\hat{\rho})$, of the parameters becomes a multivariate normal distribution. This means that it holds in general only "asymptotically", i.e., when the amount of data is "large", and only when the parameters are "far" from their limits.[17] Furthermore, the sensitivity to some of the parameters cannot disappear when $\eta = \eta_0$ (they cannot be *unidentifiable*), since then normality of the maximum likelihood estimates can then never be achieved for any amount of data. In particle physics, it is common to search for a signal of new physics in a spectrum by looking for an excess (or a deficit) of events on top of a background. This means that the above conditions are not met: the amount of data (number of events) can often be quite small, and if the signal has an unknown position and/or shape, these additional parameters are unidentifiable under the background-only hypothesis. In

[16] There are generally many different such extensions.

[17] If η_0 is on the edge of the parameter space, this can be corrected for by using a mixture of χ^2-distributions.

these cases, the only generally reliable method to approximate the distribution of Q^2 is using Monte Carlo generation of simulated data. If one requires accurate estimates of very small tail probabilities, this can require quite some computational effort.

6.6.4 Mixing up probabilities

It should be clear that the posterior probability $\Pr(H_0|\mathbf{D})$, the p-value, and the type-1 error rate α are all probabilities, but are otherwise not really related, and there is no way in which one can calculate any one of them from the others without further information. However, when a p-value is reported, this does often not stop scientists, and even more so journalists, to automatically convert the p-value into an error rate, or even more commonly a posterior probability, with potentially severe consequences. For recent examples of misreporting in the media, see, for example, the discussion in Ref. [206]. If the public receives constant reports in the media that the probability that scientists have discovered some new physics is large, say larger than 99%, but almost all these claims turn out to be false, who could blame anyone who becomes sceptic towards scientists and science, and perhaps becomes reluctant to fund science? Understanding statistics matters! Basically, the error occurs because humans naturally think about how plausible or probable different propositions are, but are ill-equipped for understanding p-values. What can happen is that an article reports that the probability that a discovery has been made is almost 1, but also that physicists are far from convinced that they actually have discovered something, two positions which of course are inconsistent. It is rather ironic that only using the frequentist notion of probability in order to be "objective" can actually make physicists appear not only very subjective, but *completely illogical*. It should be noted that also serious physicists mix up p-values, error rates, and posterior probabilities in publications, see Refs. [207,208] for recent examples – only on the topic of neutrino oscillations.

Finally, we give in Fig. 6.2 an example illustrating how conclusions using Bayesian and frequentist methods can widely disagree.

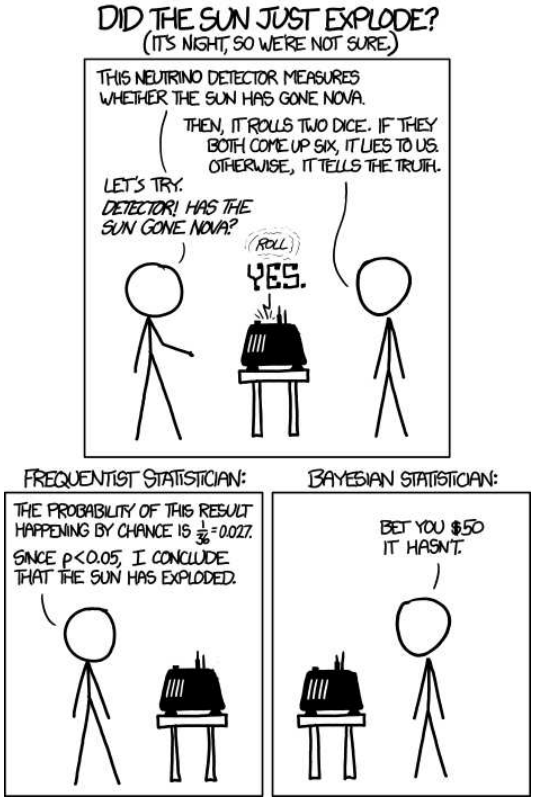

Figure 6.2. An illustration of the different conclusions which can be obtained using Bayesian and frequentist methods. http://www.xkcd.com/1132/, May 23 2013.

Chapter 7

Summary and conclusions

Part I of this thesis has dealt with the theoretical background relevant for the scientific papers presented in Part II of the thesis. The standard model of particle physics has been briefly discussed, and emphasis has been put on the topics of neutrino masses and lepton mixing, and how the resulting effects can be tested in experiments. We have discussed seesaw models, in particular the canonical type I and inverse seesaw models, and also the concepts of renormalization and renormalization group (RG) running. Different aspects of statistical inference and the associated methods have been reviewed. In Part II of the thesis, five scientific papers is presented, which investigate the models and use the techniques introduced in Part I. All the papers presented in Part II of this thesis are fundamentally motivated by the observation of neutrino oscillations, implying that neutrinos are massive and lepton flavors mixed.

In paper I [1], we have studied some very specific aspects, namely the RG running of the lepton parameters, of the inverse seesaw model, which is a high-energy model capable of accommodating neutrino masses. We have derived analytical formulas, describing the running of the neutrino parameters above the seesaw threshold in the SM and the MSSM. Also, a detailed numerical study of the RG running has been carried out. Because of the potentially large Yukawa couplings, significant running of the lepton mixing angles can be obtained. The running of the lepton mixing angles, in particular of θ_{12}, can be large if the mass spectrum of the light neutrinos is nearly degenerate. In addition, the effects of the seesaw thresholds are discussed. Some phenomenologically and theoretically interesting lepton mixing patterns, the bimaximal and tri-bimaximal patterns, can be achieved at a high-energy scale once the RG running is taken into account. Finally, the RG evolution of the light neutrino masses and of the CP-violating phases have been studied.

In paper II [2], the RG running of the neutrino parameters in a different seesaw model, the low-scale type I seesaw model with non-degenerate heavy neutrinos, has been investigated. We have shown that significant radiative corrections can be obtained at low energies, and for a short distance of RG running, as a result of

threshold effects. Analytical formulas for the RG corrections to the neutrino parameters in crossing the seesaw thresholds have been presented, indicating that the mismatch between different contributions to the mass matrix of the light neutrinos can lead to large corrections to the lepton mixing matrix. A numerical example has also been given to show that, in the presence of low-scale right-handed neutrinos, the bi-maximal mixing pattern at the TeV scale is fully compatible with the current measurements lepton mixing angles.

Massive neutrinos are the only known fermions capable of being Majorana particles, i.e., their own antiparticles, and the Majorana nature of neutrinos is intimately connected to their masses. Neutrinos being Majorana particles is equivalent to the existence of neutrinoless double beta decay, a decay which is also closely related to neutrino masses. There are, however, also other possible sources of neutrinoless double beta decay, which need to be compared with and disentangled from a pure neutrino-mediated decay. In paper III [3], we have investigated the possible future bounds on the strength of different short-range contributions to neutrinoless double beta decay. These bounds depend on the outcome of ongoing and planned experiments related to neutrino masses. Three scenarios, A, B, and C, are studied, corresponding to different combinations of experimental results. For two of the scenarios, we have determined the bounds on the coefficients ϵ_i of each point-like operator that could contribute to the decay, will for the remaining scenario, we obtain non-zero estimates.

The neutrino oscillations observed until recently correspond to the dominant effective two-flavor oscillation modes, driven by two mass-squared differences and two relatively large mixing angles, while the third angle, θ_{13}, was allowed by the data to vanish. Although the evidence for a non-zero value of the mixing angle θ_{13} was one of the most discussed topics in neutrino physics for a number of years, there had been no discussion on the standard Bayesian method to evaluate this. Reference [209] looked at the Bayesian constraints on θ_{13}, but used none of the recent data. More importantly, it only considered the constraints *assuming* θ_{13} *was non-zero*, and hence did not attempt to test whether it was non-zero or not. In paper IV [4], I described how to apply Bayesian model comparison to neutrino oscillation data, considering the evidence of a non-zero θ_{13} and the related question of whether there is evidence for CP-violation in neutrino oscillations. An interesting observation is that if no particular symmetry of the effective neutrino mass matrix is assumed, the invariance under basis redefinitions essentially fixes the prior distributions of the mixing parameters, making the Bayesian analysis unusually robust with minor prior dependence. It could be worth noting that the definite observation of a non-zero θ_{13} by Daya Bay was published while I was in the last stages of finishing the paper – without this data the evidence for a non-zero θ_{13} is much less significant. The analysis could and should be extended to compare maximal θ_{23} with θ_{23} in the two different octants, compare the different mass orderings, and use a more accurate likelihood.

Paper V [5] also deals with neutrinoless double beta decay. I performed a Bayesian global fit of the most relevant neutrinoless double beta decay experiments

within the standard model with massive Majorana neutrinos. A Bayesian analysis makes it possible to include the theoretical uncertainties on the nuclear matrix elements in a coherent manner. In addition to performing model selection and parameter estimation on the background and signal hypotheses, the aim was to evaluate the compatibility of the data used to claim the observation of neutrinoless double beta decay in Germanium with recent data using Xenon. I find moderate to strong evidence against consistency, with quite small dependence on the prior on the neutrino mass, from which the the signal rate is derived, and the uncertainties of the nuclear matrix elements.

For more detailed conclusions, the reader is referred to the corresponding papers.

Appendix A

Renormalization group equations in the type I seesaw model

In this appendix, the RGEs of the parameters of the type I seesaw model with the SM and the MSSM as underlying theories are given. The RGEs for the inverse seesaw model can be calculated as a special case of this model with six right-handed neutrinos.

A.1 SM with right-handed neutrinos

The renormalization group evolution of the parameters of the SM and the coefficient of the Weinberg operator are given by [137–140, 156, 157, 210]

$$16\pi^2\mu\frac{\mathrm{d}g_1}{\mathrm{d}\mu} = b_1 g_1^3, \tag{A.1a}$$

$$16\pi^2\mu\frac{\mathrm{d}g_2}{\mathrm{d}\mu} = b_2 g_2^3, \tag{A.1b}$$

$$16\pi^2\mu\frac{\mathrm{d}g_3}{\mathrm{d}\mu} = b_3 g_3^3, \tag{A.1c}$$

$$16\pi^2\mu\frac{\mathrm{d}Y_u}{\mathrm{d}\mu} = \left(\alpha_u + C_u^u H_u + C_u^d H_d\right) Y_u, \tag{A.1d}$$

$$16\pi^2\mu\frac{\mathrm{d}Y_d}{\mathrm{d}\mu} = \left(\alpha_d + C_d^u H_u + C_d^d H_d\right) Y_d, \tag{A.1e}$$

$$16\pi^2\mu\frac{\mathrm{d}Y_e}{\mathrm{d}\mu} = \left(\alpha_e + C_e^e H_e + C_e^\nu H_\nu^{(n)}\right) Y_e, \tag{A.1f}$$

$$16\pi^2\mu\frac{\mathrm{d}Y_\nu^{(n)}}{\mathrm{d}\mu} = \left(\alpha_\nu + C_\nu^e H_e + C_\nu^\nu H_\nu^{(n)}\right) Y_\nu^{(n)}, \tag{A.1g}$$

$$16\pi^2\mu\frac{\mathrm{d}M_{\mathrm{R}}^{(n)}}{\mathrm{d}\mu} = C_{\mathrm{R}} M_{\mathrm{R}}^{(n)} \left(Y_\nu^{(n)\dagger} Y_\nu^{(n)}\right) + C_{\mathrm{R}} \left(Y_\nu^{(n)\dagger} Y_\nu^{(n)}\right)^T M_{\mathrm{R}}^{(n)}, \tag{A.1h}$$

$$16\pi^2\mu\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \alpha_\lambda^\lambda + \alpha_\lambda^g + \alpha_\lambda^Y, \tag{A.1i}$$

$$16\pi^2\mu\frac{\mathrm{d}\kappa^{(n)}}{\mathrm{d}\mu} = \alpha_\kappa \kappa^{(n)} + \left(C_\kappa^e H_e + C_\kappa^\nu H_\nu^{(n)}\right)\kappa^{(n)} + \kappa^{(n)}\left(C_\kappa^e H_e + C_\kappa^\nu H_\nu^{(n)}\right)^T, \tag{A.1j}$$

where $H_f = Y_f Y_f^\dagger$ for $f = e, \nu, u, d$, and (n) labels the quantities relevant for the effective theory between the n-th and $(n-1)$-th thresholds. The matching between the effective theories is described in the main text. GUT charge normalization for g_1 is used, which means that g_1 is related to the conventional SM coupling $\tilde{g}_1$ as $g_1^2 = \frac{5}{3}\tilde{g}_1^2$. The coefficients determining the evolution of the gauge couplings are

$$b_1 = \frac{41}{10}, \quad b_2 = -\frac{19}{16}, \quad b_3 = -7. \tag{A.2}$$

The beta functions for the Yukawa couplings each consist of a flavor diagonal part and a flavor non-diagonal part. The flavor diagonal parts are given by

$$\alpha_u = \mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right) - \frac{17}{20}g_1^2 - \frac{9}{4}g_2^2 - 8g_3^2, \tag{A.3}$$

$$\alpha_d = \mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right) - \frac{1}{4}g_1^2 - \frac{9}{4}g_2^2 - 8g_3^2, \tag{A.4}$$

$$\alpha_e = \mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right) - \frac{9}{4}g_1^2 - \frac{9}{4}g_2^2, \tag{A.5}$$

$$\alpha_\nu = \mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right) - \frac{9}{20}g_1^2 - \frac{9}{4}g_2^2, \tag{A.6}$$

$$\alpha_\kappa = 2\mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right) + \lambda - 3g_2^2, \tag{A.7}$$

while the coefficients determining the flavor non-diagonal parts are given by

$$C_u^u = C_d^d = C_e^e = C_\nu^\nu = \frac{3}{2}, \tag{A.8}$$

$$C_u^d = C_d^u = C_e^\nu = C_\nu^e = -\frac{3}{2}, \tag{A.9}$$

$$C_{\mathrm{R}} = 1, \tag{A.10}$$

$$C_\kappa^e = -\frac{3}{2}, \quad C_\kappa^\nu = \frac{1}{2}. \tag{A.11}$$

Finally, the RGE evolution of the Higgs self-coupling constant is determined by[1]

$$\alpha_\lambda^\lambda = 6\lambda^2 - 3\lambda\left(\frac{3}{5}g_1^2 + 3g_2^2\right) + \lambda \mathrm{tr}\left(3H_u + 3H_d + H_e + H_\nu^{(n)}\right), \tag{A.12}$$

$$\alpha_\lambda^g = 3g_2^4 + \frac{3}{2}\left(\frac{3}{5}g_1^2 + 3g_2^2\right)^2, \tag{A.13}$$

$$\alpha_\lambda^Y = -8\mathrm{tr}\left(3H_u^2 + 3H_d^2 + H_e^2 + \left(H_\nu^{(n)}\right)^2\right). \tag{A.14}$$

A.2 MSSM with right-handed neutrinos

If instead the MSSM is the underlying theory, the RGEs in Eqs. (A.1) (except for Eq. (A.1i), since the parameter λ is absent in the MSSM) still hold above the supersymmetry-breaking scale, but with different coefficients. Below the scale of supersymmetry-breaking, one recovers the SM as an effective theory, and the corresponding RGEs should be used. The coefficients determining the evolution of the gauge couplings are

$$b_1 = \frac{33}{5}, \quad b_2 = 1, \quad b_3 = -3. \tag{A.15}$$

The flavor diagonal terms read

$$\alpha_u = \mathrm{tr}\left(3H_u + H_\nu^{(n)}\right) - \frac{13}{15}g_1^2 - 3g_2^2 - \frac{16}{3}g_3^2, \tag{A.16}$$

$$\alpha_d = \mathrm{tr}\left(3H_d + H_e\right) - \frac{7}{15}g_1^2 - 3g_2^2 - \frac{16}{3}g_3^2, \tag{A.17}$$

$$\alpha_e = \mathrm{tr}\left(3H_d + H_e\right) - \frac{9}{5}g_1^2 - 3g_2^2, \tag{A.18}$$

$$\alpha_\nu = \mathrm{tr}\left(3H_u + H_\nu^{(n)}\right) - \frac{3}{5}g_1^2 - 3g_2^2, \tag{A.19}$$

$$\alpha_\kappa = 2\alpha_\nu, \tag{A.20}$$

while the flavor non-diagonal are determined by

$$C_u^u = C_d^d = C_e^e = C_\nu^\nu = 3, \tag{A.21}$$

$$C_u^d = C_d^u = C_e^\nu = C_\nu^e = 1, \tag{A.22}$$

$$C_\mathrm{R} = 2, \tag{A.23}$$

$$C_\kappa^e = C_\kappa^\nu = 1. \tag{A.24}$$

The fact that $\alpha_\kappa = 2\alpha_\nu$ leads to the absence of threshold effects in the MSSM, as discussed in the main text.

[1] The interaction term is $(\lambda/4)\,\phi^4$.

Bibliography

[1] J. Bergström *et al.*, *Renormalization group running of neutrino parameters in the inverse seesaw model*, Phys. Rev. **D81**, 116006 (2010), 1004.4628.

[2] J. Bergström, T. Ohlsson and H. Zhang, *Threshold effects on renormalization group running of neutrino parameters in the low-scale seesaw model*, Phys. Lett. **B698**, 297 (2011), 1009.2762.

[3] J. Bergström, A. Merle and T. Ohlsson, *Constraining new physics with a positive or negative signal of neutrino-less double beta decay*, JHEP **05**, 122 (2011), 1103.3015.

[4] J. Bergström, *Bayesian evidence for non-zero θ_{13} and CP-violation in neutrino oscillations*, JHEP **1208**, 163 (2012), 1205.4404.

[5] J. Bergström, *Combining and comparing neutrinoless double beta decay experiments using different nuclei*, JHEP **1302**, 093 (2013), 1212.4484.

[6] J. Bergström and T. Ohlsson, *Unparticle self-interactions at the Large Hadron Collider*, Phys. Rev. **D80**, 115014 (2009), 0909.2213.

[7] Particle Data Group, J. Beringer *et al.*, *Review of Particle Physics (RPP)*, Phys. Rev. **D86**, 010001 (2012).

[8] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory* (Addison-Wesley, 1995).

[9] F. Mandl and G. Shaw, *Quantum field theory*, 2 ed. (Wiley, 2010).

[10] S. F. Novaes, *Standard model: an introduction*, (1999), hep-ph/0001283.

[11] A. Pich, *The standard model of electroweak interactions*, (2007), 0705.4264.

[12] C. Giunti and C. W. Kim, *Fundamentals of neutrino physics and astrophysics* (Oxford, 2007).

[13] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19**, 1264 (1967).

[14] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22**, 579 (1961).

[15] A. Salam, Proceedings of the 8th Nobel Symposium "Elementary Particle Theory: Reletivistic Groups and Analycity", edited by N. Svartholm, 1969.

[16] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12**, 132 (1964).

[17] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13**, 508 (1964).

[18] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145**, 1156 (1966).

[19] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13**, 321 (1964).

[20] G. Guralnik, C. Hagen and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964).

[21] T. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. **155**, 1554 (1967).

[22] ATLAS Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716**, 1 (2012), `1207.7214`.

[23] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716**, 30 (2012), `1207.7235`.

[24] A. Pich, *Effective field theory*, (1998), `hep-ph/9806303`.

[25] A. V. Manohar, Lecture Notes in Physics: Perturbative and Nonperturbative Aspects of Quantum Field Theory, edited by H. Latal and W. Schweiger, Springer, 1997, `hep-ph/9606222`.

[26] H. Georgi, *Effective field theory*, Ann. Rev. Nucl. Part. Sci. **43**, 209 (1993).

[27] D. B. Kaplan, *Five lectures on effective field theory*, (2005), `nucl-th/0510023`.

[28] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10**, 531 (1963).

[29] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49**, 652 (1973).

[30] S. Weinberg, *Baryon and lepton nonconserving processes*, Phys. Rev. Lett. **43**, 1566 (1979).

[31] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28**, 870 (1962).

[32] B. Pontecorvo, *Neutrino experiments and the problem of conservation of leptonic charge*, Sov. Phys. JETP **26**, 984 (1968).

[33] V. Gribov and B. Pontecorvo, *Neutrino astronomy and lepton charge*, Phys. Lett. **B28**, 493 (1969).

[34] F. Bonnet *et al.*, *Neutrino masses from higher than $d = 5$ effective operators*, JHEP **10**, 076 (2009), 0907.3143.

[35] S. Eliezer and A. R. Swift, *Experimental consequences of ν_e-ν_μ mixing in neutrino beams*, Nucl. Phys. **B105**, 45 (1976).

[36] H. Fritzsch and P. Minkowski, *Vector-like weak currents, massive neutrinos, and neutrino beam oscillations*, Phys. Lett. **B62**, 72 (1976).

[37] S. M. Bilenky and B. Pontecorvo, *Lepton mixing and neutrino oscillations*, Phys. Rept. **41**, 225 (1978).

[38] E. K. Akhmedov, *Neutrino physics*, p. 103 (1999), hep-ph/0001264.

[39] SAGE collaboration, J. N. Abdurashitov *et al.*, *Measurement of the solar neutrino capture rate by the Russian-American gallium solar neutrino experiment during one half of the 22-year cycle of solar activity*, J. Exp. Theor. Phys. **95**, 181 (2002), astro-ph/0204245.

[40] GNO collaboration, M. Altmann *et al.*, *Complete results for five years of GNO solar neutrino observations*, Phys. Lett. **B616**, 174 (2005), hep-ex/0504037.

[41] SNO collaboration, B. Aharmim *et al.*, *Determination of the ν_e and total ^8B solar neutrino fluxes using the Sudbury Neutrino Observatory Phase I data set*, Phys. Rev. **C75**, 045502 (2007), nucl-ex/0610020.

[42] Borexino collaboration, C. Arpesella *et al.*, *First real time detection of ^7Be solar neutrinos by Borexino*, Phys. Lett. **B658**, 101 (2008), 0708.2251.

[43] Super-Kamiokande collaboration, Y. Ashie *et al.*, *A measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I*, Phys. Rev. **D71**, 112005 (2005), hep-ex/0501064.

[44] Super-Kamiokande Collaboration, R. Wendell *et al.*, *Atmospheric neutrino oscillation analysis with sub-leading effects in Super-Kamiokande I, II, and III*, Phys. Rev. **D81**, 092004 (2010), 1002.3471.

[45] KamLAND collaboration, S. Abe *et al.*, *Precision measurement of neutrino oscillation parameters with KamLAND*, Phys. Rev. Lett. **100**, 221803 (2008), 0801.4589.

[46] K2K collaboration, M. H. Ahn *et al.*, *Measurement of neutrino oscillation by the K2K experiment*, Phys. Rev. **D74**, 072003 (2006), hep-ex/0606032.

[47] MINOS collaboration, P. Adamson *et al.*, *A study of muon neutrino disappearance using the Fermilab Main Injector neutrino beam*, Phys. Rev. **D77**, 072002 (2008), 0711.0769.

[48] MINOS collaboration, P. Adamson *et al.*, *Measurement of the neutrino mass splitting and flavor mixing by MINOS*, Phys. Rev. Lett. **106**, 181801 (2011), 1103.0340.

[49] MINOS collaboration, P. Adamson *et al.*, *First direct observation of muon antineutrino disappearance*, (2011), 1104.0344.

[50] MINOS, P. Adamson *et al.*, *Improved search for muon-neutrino to electron-neutrino oscillations in MINOS*, Phys. Rev. Lett. **107**, 181802 (2011), 1108.0015.

[51] T2K, K. Abe *et al.*, *Indication of electron neutrino appearance from an accelerator-produced off-axis muon neutrino beam*, Phys. Rev. Lett. **107**, 041801 (2011), 1106.2822.

[52] T2K Collaboration, K. Abe *et al.*, *Evidence of electron neutrino appearance in a muon neutrino beam*, (2013), 1304.0841.

[53] Double Chooz, Y. Abe *et al.*, *Indication of reactor $\overline{\nu}_e$ disappearance in the Double Chooz experiment*, Phys. Rev. Lett. **108**, 131801 (2012), 1112.6353.

[54] Daya Bay, F. An *et al.*, *Observation of electron-antineutrino disappearance at Daya Bay*, Phys. Rev. Lett. **108**, 171803 (2012), 1203.1669.

[55] Daya Bay Collaboration, F. An *et al.*, *Improved Measurement of Electron Antineutrino Disappearance at Daya Bay*, Chin. Phys. **C37**, 011001 (2013), 1210.6327.

[56] RENO, J. Ahn *et al.*, *Observation of reactor electron antineutrino disappearance in the RENO experiment*, Phys. Rev. Lett. **108**, 191802 (2012), 1204.0626.

[57] Double Chooz Collaboration, Y. Abe *et al.*, *Reactor electron antineutrino disappearance in the Double Chooz experiment*, Phys. Rev. **D86**, 052008 (2012), 1207.6632.

[58] Double Chooz Collaboration, Y. Abe *et al.*, *First Measurement of θ_{13} from Delayed Neutron Capture on Hydrogen in the Double Chooz Experiment*, (2013), 1301.2948.

[59] M. Mezzetto and T. Schwetz, *$\theta_1 3$: Phenomenology, present status and prospect*, J. Phys. **G37**, 103001 (2010), 1003.5800.

[60] M. Gonzalez-Garcia *et al.*, *Global fit to three neutrino mixing: critical look at present precision*, JHEP **1212**, 123 (2012), 1209.3023.

[61] LSND collaboration, C. Athanassopoulos *et al.*, *Results on $\nu_\mu \to \nu_e$ neutrino oscillations from the LSND experiment*, Phys. Rev. Lett. **81**, 1774 (1998), nucl-ex/9709006.

[62] LSND collaboration, A. Aguilar *et al.*, *Evidence for neutrino oscillations from the observation of $\overline{\nu_e}$ appearance in a $\overline{\nu_\mu}$ beam*, Phys. Rev. **D64**, 112007 (2001), hep-ex/0104049.

[63] MiniBooNE collaboration, A. A. Aguilar-Arevalo *et al.*, *Event excess in the MiniBooNE search for $\overline{\nu_\mu} \to \overline{\nu_e}$ oscillations*, Phys. Rev. Lett. **105**, 181801 (2010), 1007.1150.

[64] J. Kopp *et al.*, *Sterile neutrino oscillations: The global picture*, (2013), 1303.3011.

[65] V. Aseev *et al.*, *Upper limit on the electron antineutrino mass from the Troitsk experiment*, Phys. Rev. **D84**, 112003 (2011), 1108.5034.

[66] Mainz collaboration, C. Kraus *et al.*, *Final results from phase II of the Mainz neutrino mass search in tritium β decay*, Eur. Phys. J. **C40**, 447 (2005), hep-ex/0412056.

[67] MARE collaboration, A. Nucciotti *et al.*, *Neutrino mass calorimetric searches in the MARE experiment*, (2010), 1012.2290.

[68] KATRIN collaboration, A. Osipowicz *et al.*, *KATRIN: A next generation tritium beta decay experiment with sub-eV sensitivity for the electron neutrino mass*, (2001), hep-ex/0109033.

[69] KATRIN Collaboration, J. Bonn, *The status of KATRIN*, Prog. Part. Nucl. Phys. **64**, 285 (2010).

[70] S. Hannestad, *Primordial neutrinos*, Ann. Rev. Nucl. Part. Sci. **56**, 137 (2006), hep-ph/0602058.

[71] S. Hannestad and Y. Y. Y. Wong, *Neutrino mass from future high redshift galaxy surveys: Sensitivity and detection threshold*, JCAP **0707**, 004 (2007), astro-ph/0703031.

[72] Z. Hou *et al.*, *Constraints on cosmology from the cosmic microwave background power spectrum of the 2500-square degree SPT-SZ survey*, (2012), 1212.6267.

[73] J. L. Sievers *et al.*, *The Atacama Cosmology Telescope: Cosmological parameters from three seasons of data*, (2013), 1301.0824.

[74] Planck Collaboration, P. Ade *et al.*, *Planck 2013 results. XVI. Cosmological parameters*, (2013), 1303.5076.

[75] M. Doi, T. Kotani and E. Takasugi, *Double beta Decay and Majorana Neutrino*, Prog. Theor. Phys. Suppl. **83**, 1 (1985).

[76] S. M. Bilenky and S. T. Petcov, *Massive neutrinos and neutrino oscillations*, Rev. Mod. Phys. **59**, 671 (1987).

[77] H. Päs *et al.*, *Towards a superformula for neutrinoless double beta decay*, Phys. Lett. **B453**, 194 (1999), hep-ph/9804374.

[78] A. Faessler *et al.*, *Quasiparticle random phase approximation uncertainties and their correlations in the analysis of $0\nu\beta\beta$ decay*, Phys. Rev. **D79**, 053001 (2009), 0810.5733.

[79] F. Simkovic *et al.*, *$0\nu\beta\beta$-decay nuclear matrix elements with self-consistent short-range correlations*, Phys. Rev. **C79**, 055501 (2009), 0902.0331.

[80] J. Suhonen and M. Kortelainen, *Nuclear matrix elements for double beta decay*, Int. J. Mod. Phys. **E17**, 1 (2008).

[81] J. Schechter and J. W. F. Valle, *Neutrino decay and spontaneous violation of lepton number*, Phys. Rev. **D25**, 774 (1982).

[82] H. V. Klapdor-Kleingrothaus, A. Dietz and I. V. Krivosheina, *Neutrinoless double beta decay: Status of evidence*, Found. Phys. **32**, 1181 (2002), hep-ph/0302248.

[83] H. Klapdor-Kleingrothaus *et al.*, *Search for neutrinoless double beta decay with enriched ^{76}Ge in Gran Sasso 1990-2003*, Phys. Lett. **B586**, 198 (2004), hep-ph/0404088.

[84] H. Klapdor-Kleingrothaus and I. Krivosheina, *The evidence for the observation of $0\nu\beta\beta$ decay: The identification of $0\nu\beta\beta$ events from the full spectra*, Mod. Phys. Lett. **A21**, 1547 (2006).

[85] I. Kirpichnikov, *Klapdor's claim for the observation of the neutrinoless double beta-decay of Ge-76. Analysis and corrections*, (2010), 1006.2025.

[86] EXO, M. Auger *et al.*, *Search for Neutrinoless Double-Beta Decay in ^{136}Xe with EXO-200*, Phys. Rev. Lett. **109**, 032505 (2012), 1205.5608.

[87] KamLAND-Zen, A. Gando *et al.*, *Limit on neutrinoless $\beta\beta$ decay of ^{136}Xe from the first phase of KamLAND-Zen and comparison with the positive claim in ^{76}Ge*, (2012), 1211.3863.

[88] K. H. Ackermann *et al.*, *The Gerda experiment for the search of 0 $\nu\beta\beta$ decay in ^{76}Ge*, European Physical Journal C **73**, 2330 (2013), 1212.4067.

[89] GERDA, M. Agostini *et al.*, *Results on neutrinoless double beta decay of ^{76}Ge from GERDA Phase I*, (2013), 1307.4720.

[90] B. C. Allanach, C. H. Kom and H. Päs, *Large Hadron Collider probe of supersymmetric neutrinoless double beta decay mechanism*, Phys. Rev. Lett. **103**, 091801 (2009), 0902.4697.

[91] A. Faessler *et al.*, *Uncovering multiple CP-nonconserving mechanisms of $(\beta\beta)_{0\nu}$-decay*, (2011), 1103.2434.

[92] A. Faessler *et al.*, *Multi-isotope degeneracy of neutrinoless double beta decay mechanisms in the quasi-particle random phase approximation*, (2011), 1103.2504.

[93] M. Chemtob, *Phenomenological constraints on broken R parity symmetry in supersymmetry models*, Prog. Part. Nucl. Phys. **54**, 71 (2005), hep-ph/0406029.

[94] V. Tello *et al.*, *Left-right Symmetry: from LHC to neutrinoless double beta decay*, Phys. Rev. Lett. **106**, 151801 (2011), 1011.3522.

[95] F. Deppisch and H. Päs, *Pinning down the mechanism of neutrinoless double beta decay with measurements in different nuclei*, Phys. Rev. Lett. **98**, 232501 (2007), hep-ph/0612165.

[96] V. M. Gehman and S. R. Elliott, *Multiple-isotope comparison for determining $0\nu\beta\beta$ mechanisms*, J. Phys. **G34**, 667 (2007), hep-ph/0701099.

[97] H. Päs *et al.*, *A superformula for neutrinoless double beta decay. II: The short range part*, Phys. Lett. **B498**, 35 (2001), hep-ph/0008182.

[98] G. Prezeau, M. Ramsey-Musolf and P. Vogel, *Neutrinoless double beta decay and effective field theory*, Phys. Rev. **D68**, 034016 (2003), hep-ph/0303205.

[99] E. Ma, *Pathways to naturally small neutrino masses*, Phys. Rev. Lett. **81**, 1171 (1998), hep-ph/9805219.

[100] P. Minkowski, *$\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?*, Phys. Lett. **B67**, 421 (1977).

[101] T. Yanagida, In proceedings of "Workshop on the Baryon Number of the Universe and Unified Theories", edited by O. Sawada and A. Sugamoto, p. 95, 1979.

[102] M. Gell-Mann, P. Ramond and R. Slansky, Supergravity, edited by P. van Nieuwenhuizen and D. Freedman, p. 315, 1979.

[103] R. N. Mohapatra and G. Senjanović, *Neutrino mass and spontaneous parity nonconservation*, Phys. Rev. Lett. **44**, 912 (1980).

[104] M. Magg and C. Wetterich, *Neutrino mass problem and gauge hierarchy*, Phys. Lett. **B94**, 61 (1980).

[105] J. Schechter and J. W. F. Valle, *Neutrino masses in $SU(2) \times U(1)$ theories*, Phys. Rev. **D22**, 2227 (1980).

[106] T. P. Cheng and L. F. Li, *Neutrino masses, mixings and oscillations in $SU(2) \times U(1)$ models of electroweak interactions*, Phys. Rev. **D22**, 2860 (1980).

[107] C. Wetterich, *Neutrino masses and the scale of $B - L$ violation*, Nucl. Phys. **B187**, 343 (1981).

[108] G. Lazarides, Q. Shafi and C. Wetterich, *Proton lifetime and fermion masses in an $SO(10)$ model*, Nucl. Phys. **B181**, 287 (1981).

[109] R. N. Mohapatra and G. Senjanović, *Neutrino masses and mixings in gauge models with spontaneous parity violation*, Phys. Rev. **D23**, 165 (1981).

[110] R. Foot *et al.*, *Seesaw neutrino masses induced by a triplet of leptons*, Z. Phys. **C44**, 441 (1989).

[111] M. Fukugita and T. Yanagida, *Baryogenesis without grand unification*, Phys. Lett. **B174**, 45 (1986).

[112] M. Luty, *Baryogenesis via leptogenesis*, Phys. Rev. **D45**, 455 (1992).

[113] W. Buchmüller and M. Plümacher, *Baryon asymmetry and neutrino mixing*, Phys. Lett. **B389**, 73 (1996), `hep-ph/9608308`.

[114] E. Ma and U. Sarkar, *Neutrino masses and leptogenesis with heavy Higgs triplets*, Phys. Rev. Lett. **80**, 5716 (1998), `hep-ph/9802445`.

[115] S. Davidson and A. Ibarra, *A lower bound on the right-handed neutrino mass from leptogenesis*, Phys. Lett. **B535**, 25 (2002), `hep-ph/0202239`.

[116] A. Pilaftsis and T. E. Underwood, *Resonant leptogenesis*, Nucl. Phys. **B692**, 303 (2004), `hep-ph/0309342`.

[117] W. Buchmüller, P. Di Bari and M. Plümacher, *Leptogenesis for pedestrians*, Ann. Phys. **315**, 305 (2005), hep-ph/0401240.

[118] F. del Aguila, J. Aguilar-Saavedra and R. Pittau, *Heavy neutrino signals at large hadron colliders*, JHEP **0710**, 047 (2007), hep-ph/0703261.

[119] P. Nath *et al.*, *The hunt for new physics at the Large Hadron Collider*, Nucl. Phys. Proc. Suppl. **200-202**, 185 (2010), 1001.2693.

[120] A. Broncano, M. Gavela and E. E. Jenkins, *The effective Lagrangian for the seesaw model of neutrino mass and leptogenesis*, Phys. Lett. **B552**, 177 (2003), hep-ph/0210271.

[121] A. Broncano, M. Gavela and E. E. Jenkins, *Neutrino physics in the seesaw model*, Nucl. Phys. **B672**, 163 (2003), hep-ph/0307058.

[122] A. Abada *et al.*, *Low energy effects of neutrino masses*, JHEP **0712**, 061 (2007), 0707.4058.

[123] M. Gavela *et al.*, *Minimal flavour seesaw models*, JHEP **0909**, 038 (2009), 0906.1461.

[124] Neutrino Factory/Muon Collider collaboration, C. H. Albright *et al.*, *The neutrino factory and beta beam experiments and development*, (2004), physics/0411123.

[125] ISS Physics Working Group, A. Bandyopadhyay *et al.*, *Physics at a future Neutrino Factory and super-beam facility*, Rept. Prog. Phys. **72**, 106201 (2009), 0710.4947.

[126] ISS Detector Working Group, T. Abe *et al.*, *Detectors and flux instrumentation for future neutrino facilities*, JINST **4**, T05001 (2009), 0712.4129.

[127] ISS Accelerator Working Group, M. Apollonio *et al.*, *Accelerator design concept for future neutrino facilities*, JINST **4**, P07001 (2009), 0802.4023.

[128] *The international design study for the neutrino factory*, http://www.ids-nf.org.

[129] S. Antusch, J. P. Baumann and E. Fernandez-Martinez, *Non-standard neutrino interactions with matter from physics beyond the standard model*, Nucl. Phys. **B810**, 369 (2009), 0807.1003.

[130] S. Davidson, *Parametrizations of the seesaw, or, can the seesaw be tested?*, (2004), hep-ph/0409339.

[131] J. A. Casas and A. Ibarra, *Oscillating neutrinos and* $\mu \to e, \gamma$, Nucl. Phys. **B618**, 171 (2001), hep-ph/0103065.

[132] R. Mohapatra and J. Valle, *Neutrino mass and baryon number nonconservation in superstring models*, Phys. Rev. **D34**, 1642 (1986).

[133] M. Farina, D. Pappadopulo and A. Strumia, *A modified naturalness principle and its experimental tests*, (2013), 1303.7244.

[134] S. P. Martin, *A supersymmetry primer*, (1997), hep-ph/9709356.

[135] H. Baer and X. Tata, *Weak scale supersymmetry: From superfields to scattering events* (Cambridge University Press, 2006).

[136] I. Aitchison, *Supersymmetry in particle physics. An elementary introduction* (Cambridge University Press, 2007).

[137] P. H. Chankowski and Z. Pluciennik, *Renormalization group equations for seesaw neutrino masses*, Phys. Lett. **B316**, 312 (1993), hep-ph/9306333.

[138] K. S. Babu, C. N. Leung and J. T. Pantaleone, *Renormalization of the neutrino mass operator*, Phys. Lett. **B319**, 191 (1993), hep-ph/9309223.

[139] S. Antusch *et al.*, *Neutrino mass operator renormalization revisited*, Phys. Lett. **B519**, 238 (2001), hep-ph/0108005.

[140] S. Antusch *et al.*, *Neutrino mass operator renormalization in two Higgs doublet models and the MSSM*, Phys. Lett. **B525**, 130 (2002), hep-ph/0110366.

[141] W. Chao and H. Zhang, *One-loop renormalization group equations of the neutrino mass matrix in the triplet seesaw model*, Phys. Rev. **D75**, 033003 (2007), hep-ph/0611323.

[142] M. A. Schmidt, *Renormalization group evolution in the type I + II seesaw model*, Phys. Rev. **D76**, 073010 (2007), 0705.3841.

[143] J. Chakrabortty *et al.*, *Renormalization group evolution of neutrino masses and mixing in the Type-III seesaw mechanism*, Nucl. Phys. **B820**, 116 (2009), 0812.2776.

[144] V. D. Barger *et al.*, *Bimaximal mixing of three neutrinos*, Phys. Lett. **B437**, 107 (1998), hep-ph/9806387.

[145] D. V. Ahluwalia, *On reconciling atmospheric, LSND, and solar neutrino oscillation data*, Mod. Phys. Lett. **A13**, 2249 (1998), hep-ph/9807267.

[146] H. Fritzsch and Z. z. Xing, *Large leptonic flavor mixing and the mass spectrum of leptons*, Phys. Lett. **B440**, 313 (1998), hep-ph/9808272.

[147] A. J. Baltz, A. S. Goldhaber and M. Goldhaber, *The Solar neutrino puzzle: An oscillation solution with maximal neutrino mixing*, Phys. Rev. Lett. **81**, 5730 (1998), hep-ph/9806540.

[148] P. F. Harrison, D. H. Perkins and W. G. Scott, *Tri-bimaximal mixing and the neutrino oscillation data*, Phys. Lett. **B530**, 167 (2002), hep-ph/0202074.

[149] P. F. Harrison and W. G. Scott, *Symmetries and generalisations of tri-bimaximal neutrino mixing*, Phys. Lett. **B535**, 163 (2002), hep-ph/0203209.

[150] Z. z. Xing, *Nearly tri-bimaximal neutrino mixing and CP violation*, Phys. Lett. **B533**, 85 (2002), hep-ph/0204049.

[151] J. A. Casas *et al.*, *General RG equations for physical neutrino parameters and their phenomenological implications*, Nucl. Phys. **B573**, 652 (2000), hep-ph/9910420.

[152] P. H. Chankowski, W. Krolikowski and S. Pokorski, *Fixed points in the evolution of neutrino mixings*, Phys. Lett. **B473**, 109 (2000), hep-ph/9910231.

[153] S. Antusch *et al.*, *Running neutrino masses, mixings and CP phases: Analytical results and phenomenological consequences*, Nucl. Phys. **B674**, 401 (2003), hep-ph/0305273.

[154] S. Antusch *et al.*, *Running neutrino mass parameters in see-saw scenarios*, JHEP **03**, 024 (2005), hep-ph/0501272.

[155] J. w. Mei, *Running neutrino masses, leptonic mixing angles and CP-violating phases: From M_Z to Λ_{GUT}*, Phys. Rev. **D71**, 073012 (2005), hep-ph/0502015.

[156] S. Ray, *Renormalization group evolution of neutrino masses and mixing in seesaw models: A review*, Int. J. Mod. Phys. **A25**, 4339 (2010), 1005.1938.

[157] Y. Lin, L. Merlo and A. Paris, *Running effects on lepton mixing angles in flavour models with type I seesaw*, Nucl. Phys. **B835**, 238 (2010), 0911.3037.

[158] T. Appelquist and J. Carazzone, *Infrared singularities and massive fields*, Phys. Rev. **D11**, 2856 (1975).

[159] S. Antusch *et al.*, *Neutrino mass matrix running for non-degenerate see-saw scales*, Phys. Lett. **B538**, 87 (2002), hep-ph/0203233.

[160] R. Cox, *Probability, frequency, and reasonable expectation*, Am. J. Phys. **14**, 1 (1946).

[161] E. T. Jaynes, *Probability theory: The logic of science* (Cambridge University Press, 2003).

[162] T. J. Loredo, *From laplace to supernova SN 1987A: Bayesian inference in astrophysics*, Maximum-Entropy and Bayesian Methods , 81 (1990).

[163] E. V. Linder and R. Miquel, *Tainted evidence: Cosmological model selection vs. fitting*, Int. J. Mod. Phys. **D17**, 2315 (2008), astro-ph/0702542.

[164] A. R. Liddle *et al.*, *Comment on "Tainted evidence: Cosmological model selection versus fitting", by Eric V. Linder and Ramon Miquel (astro-ph/0702542v2)*, (2007), astro-ph/0703285.

[165] S. Nesseris and J. Garcia-Bellido, *Is the Jeffreys' scale a reliable tool for Bayesian model comparison in cosmology?*, (2012), 1210.7652.

[166] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology*, Contemp. Phys. **49**, 71 (2008), 0803.4089.

[167] R. Trotta, *Applications of Bayesian model selection to cosmological parameters*, Mon. Not. Roy. Astron. Soc. **378**, 72 (2007), astro-ph/0504022.

[168] M. Hobson *et al.* (Eds.), *Bayesian methods in cosmology* (Cambridge University Press, 2010).

[169] F. Feroz *et al.*, *Bayesian selection of sign(μ) within mSUGRA in global fits including WMAP5 results*, JHEP **0810**, 064 (2008), 0807.4512.

[170] S. S. AbdusSalam *et al.*, *Selecting a model of supersymmetry breaking mediation*, Phys. Rev. **D80**, 035017 (2009), 0906.0957.

[171] H. Jeffreys, *Theory of probability* (Oxford University PRess, 1961).

[172] R. E. Kass and A. E. Raftery, *Bayes Factors*, J. Am. Stat. Ass. **90**, 773 (1995).

[173] A. Caldwell and K. Kroninger, *Signal discovery in sparse spectra: A Bayesian analysis*, Phys. Rev. **D74**, 092003 (2006), physics/0608249.

[174] CLAS, D. Ireland *et al.*, *A Bayesian analysis of pentaquark signals from CLAS data*, Phys. Rev. Lett. **100**, 052001 (2008), 0709.3154.

[175] R. D. Cousins, *Comment on 'Bayesian analysis of pentaquark signals from CLAS data', with response to the reply by Ireland and Protopopsecu*, Phys. Rev. Lett. **101**, 029101 (2008), 0807.1330.

[176] D. G. Ireland and D. Protopopescu, *Ireland and Protopopescu Reply*, Phys. Rev. Lett. **101**, 029102 (2008).

[177] F. Beaujean *et al.*, *Bayesian fit of wxclusive $b \to s\bar{\ell}\ell$ decays: The standard model operator basis*, JHEP **1208**, 030 (2012), 1205.1838.

[178] D. Parkinson and A. R. Liddle, *Bayesian model averaging in astrophysics: A review*, Statistical Analysis and Data Mining **1**, 2013 (6), 1302.1721.

[179] N. Karpenka *et al.*, *Bayesian constraints on dark matter halo properties using gravitationally-lensed supernovae*, (2012), 1207.3708.

[180] R. E. Kass and L. Wasserman, *The selection of prior distributions by formal rules*, J. Am. Stat. Ass. **91**, 1343 (1996).

[181] D. S. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial* (Oxford University Press, 2006).

[182] J. O. Berger, *Statistical decision theory and Bayesian analysis* (Springer-Verlag, 1985).

[183] N. Haba and H. Murayama, *Anarchy and hierarchy*, Phys. Rev. **D63**, 053010 (2001), hep-ph/0009174.

[184] J. Espinosa, *Anarchy in the neutrino sector?*, (2003), hep-ph/0306019.

[185] Y. Bai and G. Torroba, *Large N (= 3) neutrinos and random matrix theory*, JHEP **1212**, 026 (2012), 1210.2394.

[186] A. de Gouvea and H. Murayama, *Statistical test of anarchy*, Phys. Lett. **B573**, 94 (2003), hep-ph/0301050.

[187] A. de Gouvea and H. Murayama, *Neutrino mixing anarchy: Alive and kicking*, (2012), 1204.1249.

[188] P. Marshall, N. Rajguru and A. Slosar, *Bayesian evidence as a tool for comparing datasets*, Phys. Rev. **D73**, 067302 (2006), astro-ph/0412535.

[189] F. Feroz *et al.*, *Are BR($\overline{B} \to X_s\gamma$) and $(g-2)_\mu$ consistent within the Constrained MSSM?*, (2009), 0903.2487.

[190] M. Hobson, S. Bridle and O. Lahav, *Combining cosmological datasets: hyperparameters and bayesian evidence*, Mon. Not. Roy. Astron. Soc. **335**, 377 (2002), astro-ph/0203259.

[191] N. Metropolis *et al.*, *Equation of state calculations by fast computing machines*, J. Chem. Phys. **21**, 1087 (1953).

[192] W. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57**, 97 (1970).

[193] J. Skilling, *Nested sampling*, AIP Conf. Proc. **735**, 395 (2004).

[194] J. Skilling, *Nested sampling for general Bayesian computation*, Bayesian analysis **1**, 833 (2006).

[195] F. Feroz and M. Hobson, *Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis*, Mon. Not. Roy. Astron. Soc. **384**, 449 (2008), 0704.3704.

[196] F. Feroz, M. Hobson and M. Bridges, MultiNest*: an efficient and robust Bayesian inference tool for cosmology and particle physics*, Mon. Not. Roy. Astron. Soc. **398**, 1601 (2009), 0809.3437.

[197] F. Feroz *et al.*, *Importance nested sampling and the* MultiNest *algorithm*, (2013), 1306.2144.

[198] E. Cameron and A. Pettitt, *Recursive pathways to marginal likelihood estimation with prior-sensitivity analysis*, (2013), 1301.6450.

[199] J. M. Marin and C. Robert, *On resolving the Savage-Dickey paradox*, Electron. J. Statist. **4**, 643 (2010), 0910.1452.

[200] A. R. Liddle, *Information criteria for astrophysical model selection*, MNRAS **377**, L74 (2007), arXiv:astro-ph/0701113.

[201] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. **57**, 3873 (1998), arXiv:physics/9711021.

[202] G. Cowan *et al.*, *Asymptotic formulae for likelihood-based tests of new physics*, European Physical Journal C **71**, 1554 (2011), 1007.1727.

[203] X. L. Meng, *Posterior predictive p-values*, Annals of Statistics **22**, 1142 (1994).

[204] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9**, 60 (1938).

[205] R. Protassov *et al.*, *Statistics: handle with care, detecting multiple model components with the likelihood ratio test*, (2002), astro-ph/0201547.

[206] G. D'Agostini, *Probably a discovery: Bad mathematics means rough scientific communication*, (2011), 1112.3620.

[207] H. Burroughs *et al.*, *Calculating error bars for neutrino mixing parameters*, Phys. Rev. **C85**, 068501 (2012), 1204.1354.

[208] E. Ciuffoli, J. Evslin and X. Zhang, *Confidence in a neutrino mass hierarchy determination*, (2013), 1305.5150.

[209] H. Ge, C. Giunti and Q. Liu, *Bayesian constraints on θ_{13} from solar and KamLAND neutrino data*, Phys. Rev. **D80**, 053009 (2009), 0810.5443.

[210] J. Kersten, *Quantum corrections to the lepton flavor structure and applications*, PhD thesis, TU München, 2004.

Part II

Scientific papers

I

Paper I

J. Bergström, M. Malinský, T. Ohlsson, and H. Zhang

Renormalization group running of neutrino parameters in the inverse seesaw model

Physical Review **D81**, 116006 (2010)

arXiv:1004.4628

II

Paper II

J. Bergström, T. Ohlsson, and H. Zhang
Threshold effects on renormalization group running of neutrino parameters in the low-scale seesaw model
Physics Letters **B698**, 297 (2011)
arXiv:1009.2762

Paper III

J. Bergström, A. Merle, and T. Ohlsson
Constraining new physics with a positive or negative signal of neutrino-less double beta decay
Journal of High Energy Physics **05**, 122 (2011)
arXiv:1103.3015

III

Paper IV

J. Bergström

Bayesian evidence for non-zero θ_{13} and CP-violation in neutrino oscillations

Journal of High Energy Physics **08**, 163 (2012)

arXiv:1205.4404

IV

Paper V

J. Bergström

Combining and comparing neutrinoless double beta decay experiments using different nuclei

Journal of High Energy Physics **02**, 093 (2013)

arXiv:1212.4484

V